DOCUMENT RESUME

ED 296 144 CE 050 431

AUTHOR Morris, Robin; And Others
TITLE Using Basic Skills Testing To Improve the Effectiveness Remediation in Employment and Training Programs for Youth. Research Report 88-05.
INSTITUTION National Commission for Employment Policy (DOL), Washington, D.C.
PUB DATE May 88
NOTE 85p.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC04 Plus Postage.
DESCRIPTORS Adult Basic Education; *Basic Skills; Criterion Referenced Tests; *Employment Potential; Employment Programs; Federal Programs; *Job Skills; Job Training; Screening Tests; Skill Development; Standardized Tests; Test Reviews; *Test Selection; *Test Use
IDENTIFIERS *Job Training Partnership Act 1982

ABSTRACT

This paper was developed to help Job Training Partnership Act (JTPA) administrators make informed decisions when selecting employability assessment tools. The paper focuses on one aspect of participant assessment: assessing the level of basic education skills of economically disadvantaged youth. The paper provides the following: (1) comparative information on some of the most widely used basic skills tests within the JTPA system--both standardized and criterion-referenced; (2) examples of how assessment data can be used to improve program planning and participant impact; and (3) policy recommendations for consideration at the state and local levels. Eighteen tests of basic skills are profiled, with information listed including publisher, norms, administration, cost, subtest areas, and reviewer's comments focused on appropriateness of the test for JTPA clients and recommendations concerning the test's best use. (KC)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED296144

USING BASIC SKILLS TESTING TO IMPROVE THE EFFECTIVENESS OF REMEDIATION IN EMPLOYMENT AND TRAINING PROGRAMS FOR YOUTH

by

Robin Morris, Ph.D.
Lori Strumpf, M.Ed., Ed.S.
Susan Curnan, M.S., M.F.S.

Frances R. Rothstein, Editor

May, 1988 RR-88-05

RESEARCH REPORT SERIES
NATIONAL COMMISSION
FOR EMPLOYMENT POLICY
1522 K STREET, N.W.
WASHINGTON, D.C. 20005

CE050431

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

- This document has been reproduced as received from the person or organization originating it.
- Minor changes have been made to improve reproduction quality.
- Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

BEST COPY AVAILABLE

This paper was produced by the Center for Remediation Design for the National Commission for Employment Policy. The Center for Remediation Design is supported by the U.S. Conference of Mayors, the National Association of Counties, the National Association of Private Industry Councils, the National Alliance of Business, and the National Job Training Partnership, Inc. The views expressed are those of the authors and do not represent the official positions of any of the individual organizational sponsors nor do they necessarily reflect the views of the National Commission for Employment Policy or its staff.

Robin Morris, Ph.D., is an Assistant Professor in the Department of Psychology and Director of the Assessment Lab at Georgia State University. Lori Strumpf, M.Ed., Ed.S., is the Director for the Center for Remediation Design in Washington, D.C. Susan Curnan, M.S., M.F.S., is the Director for the Program Assistance Center for Human Resources at the Heller Graduate School, Brandeis University in Waltham, Massachusetts.

The authors would like to express thanks to a number of individuals who contributed greatly to the development and publication of this study. First, we wish to thank Frances R. Rothstein for her editorial services. Second, we want to extend our gratitude to the following persons for their significant input and suggestions: Evelyn Ganzglass, Program Director, Employment and Training, National Governors' Association; John Hawk, Personnel Research Psychologist, United States Department of Labor; Karen West, Deputy Director, Jobs Central, Flint, Michigan; and John Haycook, Manager, Planning and Program Development, Region 7B, Employment and Training Consortium, Harrison, Michigan; and Kay F. Albright, Deputy Director, National Commission for Employment Policy, who served as project officer.

CONTENTS

CHAPTER I: Introduction 1

Policy Recommendations for States, Private Industry Councils, and Service Delivery Areas to Improve Assessment Strategies and Service Delivery 2

National Policy Recommendations 3

Recommendations for State Assessment Priorities 4

Recommendations for Local Decision Makers 5

Assumptions in Developing This Paper 6

CHAPTER II: Current Basic Skills Testing Practices 8

Survey Findings 8

Related Issues 10

CHAPTER III: Using Testing to Define Youth Intervention Needs 12

An Employability Continuum 12

The Four Assessment Steps 15

Using Assessment Strategies to Define Target Groups 16

CHAPTER IV: Overview of The Appraisal Process: Test and Data Gathering Issues 18

The Appraisal Process 18

Testing Issues Regarding the Five Basic Skills 21

Reading Comprehension 22

Math Computation 23

Written Communication 24

Verbal Communication 25

Problem-Solving 25

CHAPTER V: Test Selection and Measurement Issues 27

Defining Deficiency 27

Functions and Limitations of Testing/Assessment 29

Validity 29

Reliability 31

Individual versus Group Testing 32

Multifactorial Tests 33

Classification Errors Based on Test Data 34

Normative Data Needs, Test Bias, Special JTPA Population Needs 34

Disadvantages of Grade-Equivalent Test Scores 36

Diagnosis 37

Monitoring Progress 37

Pre- and Post-Testing 38

Delineating Job-Specific Skills 38

CHAPTER VI: Types of Tests and Their Uses 42

Formalized/Standardized Tests (Paper and Pencil 42

Intelligence Tests 43

Aptitude Tests 44

Achievement Tests 44

Personality Tests 45

Occupational Skills Tests 46

Criterion-Referenced Testing and Competency-Based Programs 46

The CASAS Example 49

Characteristics of Criterion-Referenced Tests 49

Appendix A: Test Descriptions 52

IOWA Test of Basic Skills (1985 Edition) . . 54

Tests of Achievement and Proficiency (TAP, 1985 Edition) 55

Cognitive Abilities Test 56

MULTISCORE 57

Gates-MacGinitie Reading Test 58

Woodcock-Johnson Psycho-Educational Battery . 59

Test of Written Language (TOWL) 61

Gray Oral Reading Test (GOAT) 62

Detroit Tests of Learning Aptitude 63

Wide Range Achievement Test-Revised (WRAT-R) 64

Kaufman Test of Educatinal Achievement (K-TEA) 65

Adult Basic Learning Examination (ABLE - 2nd Edition) 66

Tests of Adult Basic Education (TABE) (Forms 5 and 6) 67

USES Basic Occupational Literacy Test (BOLT) 68

Woodcock Reading Mastery Tests - Revised . . 69

Key Math Diagnostic Test 70

Peabody Individual Achievement Test 71

CASAS Adult Life Skills Pre-Employment Tests 72

Appendix B: JTPA Survey Conducted by the Center for Remediation Design with Brandeis University 73

Survey Form 73

Survey Methodology 74

Survey Findings 75

Survey Participants 78

CHAPTER I

INTRODUCTION

Policy makers at the national, state, and local level are developing policies and programs which will assist in providing basic academic skills and work skills to tomorrow's labor force. Within this context, the nation's employment and training system is being called upon to provide remedial training in work related basic academic skills to economically disadvantaged youth. The Job Training Partnership Act is one vehicle for providing remediation services to both youth and adults lacking basic academic skills. Both Titles IIA and IIB could be more fully developed to enhance remediation strategies already begun. The 1986 amendments to the Job Training Partnership Act call for providing basic skills remediation to youth who are deficient in basic skills, and who participate in the summer youth employment program. The amendments also re-focus the funds within the Act specifically earmarked for JTPA coordination with education. Those funds, commonly referred to as 8% funds, are now to be used to provide literacy training to youth and adults; dropout prevention and re-enrollment services to youth, giving priority to youth who are at risk of dropping out; and to develop statewide school-to-work transition programs.

The lack of basic academic skills among the nation's youth and their effect on the productivity of the nation has been acknowledged by Department of Labor officials as well as by the Congress. Congressional concern translated into specific legislative requirements. The Department of Labor's concern translates into encouragement and incentives to provide more basic skills remediation within JTPA. The Department of Labor is providing technical assistance in this area as well as proposing changes in the youth performance standards system to allow SDAs and PICs to focus on providing basic skills remediation services to at risk youth. Revisions to the performance standards have been proposed with several goals in mind: providing services to the hard to serve; providing more basic skills; and increasing the quality of training services.

All over the country, JTPA practitioners are struggling to plan, design and implement quality basic academic skills remediation programs for youth. There is no one model, or blueprint, for the "best" design. Whether done in collaboration with schools, with adult basic education programs, with community-based organizations, with industry or alone, one question seems to be asked most frequently -- what is the best approach for assessing basic academic skill deficiencies among JTPA youth?

The purpose of this paper is to assist the JTPA community in making informed decisions when selecting employability assessment tools. This paper focuses on one aspect of participant assessment: assessing the level of basic education skills. Selecting appropriate basic education skills assessments may be new to many JTPA practitioners.

This paper provides:

- Comparative information on some of the most widely used basic skills assessment strategies within the JTPA system -- both standardized and criterion-referenced;
- Examples of how assessment data can be used to improve program planning and participant impact; and
- Policy recommendations for consideration at the state and local level.

This paper is not intended to identify the "best test." There is no one best test. Assessment is an ongoing process and as such is as much an art as a science -- no perfect or complete strategy exists. Many variables affect the test selection aspect of the assessment process: the target groups, the participant outcomes expected, the amount and type of existing assessment information available, and the amount of dollars available. What is best for the needs of one program and client group may not be as effective for another.

This paper does sort through the labyrinth of information on assessment, presenting the information in a straightforward manner designed to assist JTPA practitioners. The information is presented in such a way as to inform the decision-making process that each SDA must go through to select an assessment strategy. The assessment strategy which meets local needs is the one that will help develop an accurate reflection of a youth's basic skill levels so that the JTPA system can provide the most appropriate set of services which teach youth the skills they need to become employable.

POLICY RECOMMENDATIONS FOR STATES, PRIVATE INDUSTRY COUNCILS, AND SERVICE DELIVERY AREAS TO IMPROVE ASSESSMENT STRATEGIES AND SERVICE DELIVERY

This paper addresses itself primarily to PIC and SDA staff. However, the information may assist in the development of state and local policies around assessment strategies.

Developing assessment strategies and selecting the appropriate tests is not an exact science. For employment and training practitioners the process of selecting an appropriate basic skills assessment test may prove to be frustrating.

For the short-term, practitioners are faced with sorting through a lot of information on test selection. This information will often lead to the selection of a standardized test -- a test which, in essence, describes an individual's skills as related to how groups of the same type of individual have performed.

For the purposes of JTPA basic skills testing, the use of standardized tests presents two problems. First, none of the standardized tests can compare a JTPA clients' score in relationship to other JTPA clients. Second, these standardized tests measure what a person knows in relationship to the basic skills, not what a person can do with that basic skills knowledge.

Increased public demand for job training accountability has reinforced the critical nature of basic skills assessment in the employability development arena. At the same time, mounting concerns in the employment and training community about cost-effective programming are challenging practitioners to build on what is known about basic skills assessment in systematic and expedient ways. Perhaps more than any other program component, client assessment of basic education skills is fundamental to cost-effective job training programming and ultimately labor force productivity.

It is the opinion of the authors that the following policy recommendations, if implemented, could move the employment and training system toward the development of relevant, employment-related, basic skills tests.

National Policy Recommendations

Four recommendations aimed at strengthening national leadership while maintaining local flexibility:

- Establish a common definition of "employability" based on basic education skills and work maturity deficiency levels, rather than on acquisition of the high school diploma. All evidence indicates that employers consistently rate basic education skills and work maturity as the most essential qualifications to get and maintain a job. Defining employability in these terms will enable states to set training priorities for youth.

- Require that JTPA youth employment competency systems provide a combination training program of basic education skills and, either pre-employment, work maturity or job specific skills, and thereby ensure that "employment competent" includes at least a locally acceptable snapshot of employment-related basic skills.

- Require SDAs to report basic skills information (at least reading level) through the management information system (MIS). Retain local flexibility in assessment strategies but encourage and allow for reporting grade level or strictly criterion- or competency-referenced assessment data. This point-in-time data can later be used to adjust national performance standards and allocate resources based on the location and degree of need.

- Develop a performance standard that measures outcomes for young people who are most at risk of remaining structurally unemployed because of their lack of both basic education skills and work maturity skills. This would enable states to provide incentive funds to SDAs which serve those individuals.

Recommendations for State Assessment Priorities

Four recommendations to improve the quality of employment preparation programs genuinely designed around employer needs and characteristics of unemployed youth:

- Acknowledge the problems inherent in the use of standardized norm-referenced tests, while recognizing the inevitability of their continued use for the near term. As a long-term strategy, move toward the development and increased use of state-wide criterion- or competency-referenced tests rather than trying to norm standardized tests on JTPA populations.

- Facilitate the development of a state "employability credential" with emphasis on basic education skills and work maturity. Establish functional competencies necessary for the client to obtain the employability credential. Recommend effective and acceptable assessment tools. Promote marketability of such a credential for entry level workers.

- Sponsor a statewide evaluation of current assessment practices to determine the employment connection; the efficiency of resource allocation; and the impact on youth employment preparation. Use evaluation results to determine common and unique qualities about SDAs. Provide intensive training and technical assistance to SDAs to assure credibility and usefulness of assessment data.

- Provide a common definition of "youth-at-risk" at least between education and employment and training institutions (see national recommendation, above). Provide incentive funds to SDAs which serve young people who are at risk because they lack both basic education skills and work maturity skills.

Recommendations for Local Decision Makers

Four recommendations to strengthen local programs:

- Start making decisions based on assessment of basic skills deficits of the youth. Use client assessment data to assign an individual to an appropriate set or level of services. At a minimum, collect a snapshot of basic education skills and work maturity and establish three levels of training as described in the body of this paper.

- Involve the local employer community in program development. Engage employers to verify priority training areas, assessment strategies, and certification of employability.

- Use assessment results to develop programs that include the following proven design principles for at-risk youth, and revise RFP guidelines as necessary to incorporate these program design principles:

  -- Programs must combine work and education

  -- Programs must provide "intensity" of training

  -- Programs must be delivered through alternative settings (other than traditional classrooms)

  -- Programs must be individualized and competency-based

  -- Programs must provide a management system that relates assessment to curriculum to instruction.

- Provide professional development and training opportunities for line staff and management staff in order to strengthen the connection between what is tested and what is thought to improve the overall quality of programs and staff. Through targeted training, stimulate a comprehensive assessment strategy which includes written and oral questioning, product review, interviews, and performance review. This type of mixed assessment strategy acknowledges the importance of and the relationship between an individual's basic skill knowledge levels and his or her ability to apply that knowledge.

These three sets of recommendations, focused on development of common definitions, sets of competencies which relate basic skills to work skills, and development of assessment instruments which assess participants' achievements or deficiencies in those competencies, will move the current system forward in many ways:

- The recommendations establish a top/down and bottom/up collaboration process between education and employment and training. They further delineate the role JTPA has in providing basic skills remediation.

- They assist PICs and SDAs in developing curricula.

- They recommend the use of criterion- and competency-referenced tests rather than tests based on grade level ratings as a way to help direct local decision-making regarding selection of "the best test" of basic skills.

- They provide a basis for mobility between labor markets within a state.

- They help the employment and training system to articulate to employers what specific, job-related academic skills the JTPA system provides and which of those skills a participant has achieved.

- Finally, they offer a cost-effective and time-saving strategy for developing information and tools that each SDA needs to provide effective work related basic skills remediation services.

ASSUMPTIONS IN DEVELOPING THIS PAPER

The authors of this paper identified several working assumptions around which this paper was developed. It is useful to review them briefly so the reader will understand the "voice" and perspective of the paper.

- JTPA does have a role in providing basic skills remediation as a program service to youth who are deficient. Employers are identifying the lack of basic skills as one reason youth (and workers in general) are not either employable or able to retain jobs. JTPA's job is to develop those skills necessary to get and keep jobs. Those basic academic skills which assist youth in getting and keeping jobs are therefore within the purview of the employment and training system.

- The focus of this paper is on assessing the basic skills of youth, primarily because when the demographics are reviewed they underscore that it is this part of the new labor force which puts the economy most at risk of noncompetitiveness.

- Assessment is not the same for all youth and strategies must be developed which can fit individual needs.

- Assessment is an ongoing process and the information gathered is used to adapt the program to the needs of the participant.

- Assessment of basic skills is not always done by using a formal, standardized paper and pencil test. However, testing is the focus of this paper, and is an important element because it is so widely used.

- Unless otherwise noted, this paper does not address assessing youth who have already been identified as specific learning disabled or mentally deficient through some other system.

- JTPA practitioners acknowledge that there is a relationship between **the way target groups are identified, assessment strategies are developed, curriculum is chosen, and instructional methodologies are delivered.** Therefore, a discussion on basic skills assessment is out of context without some discussion of these other issues. In other words, target groups are defined in a way which relates to having an academic deficiency; the assessment tools assess for those specific deficiencies; the curriculum is chosen because it will enhance and upgrade the skills identified as lacking (not an unrelated set of skills); and instruction maximizes the potential for gain.

- Finally, this paper will not make you a testing expert. Rather, it will assist in decision making on how to provide quality services to youth.

CHAPTER II

CURRENT BASIC SKILLS TESTING PRACTICES

This chapter reports the findings of a recent survey on basic skills testing practices in the JTPA system. The authors discuss:

- The system's current practices regarding basic skills testing and the ways in which testing results are incorporated into program design.
- Barriers that limit implementation of effective basic skills remediation programs.
- Related developments beyond those revealed by a survey of JTPA practitioners.

SURVEY FINDINGS

How are most JTPA programs testing for basic skills attainment now? What are the stress points and vital signs in the field? To begin to answer these questions and to help shape this paper, the Center for Remediation Design, together with Brandeis University, conducted a series of telephone interviews with JTPA affiliates during August, 1987. (Appendix B contains the entire set of questions and the distribution of the sample together with the number of responses, state by state. A total of 150 programs out of an originally randomly selected sample of 205 participated.)

Overall, the report from the field is encouraging, at times even surprising with regard to the advances made toward refining basic skills testing techniques and incorporating basic skills remediation into local programs in the absence of specific guidance or training. For example:

- Nearly 70 percent of the programs sampled provide basic skills remediation both in summer and during the school year, while 28 percent limit basic skills remediation to the summer only.
- Although most programs reported using a variety of instructional techniques, among the most impressive findings is that more than 70 percent of the programs now use computers as teaching tools, nearly 75 percent employ genuine individualized competency-based techniques and nearly 60 percent tied basic skills

instruction to work experience, thereby modeling some of the most critical elements of effective programs for at-risk youth.

- Eighty-five percent of the programs explained that basic skills remediation was a function of their JTPA youth employment competency system.

- When asked how competency gains were measured, nearly 25 percent reported using grade level advances followed closely by 21 percent reporting criterion-referenced or functional skill gains (often to supplement, rather than to replace, grade level scores).

- Others reported defining attainment through some combination of grade level scores and GED test scores.

The single most revealing question regarding both summer and year-round testing practices was: "What tests(s) do you use?" The 92 percent of respondents who reported administering standardized tests most commonly used the following:

Tests of Adult Basic Education (TABE): Used by more than 39 percent of programs

California Achievement Test (CAT): Used by more than 22 percent of programs

Wide-Range Achievement Test (WRAT): Used by nearly 17 percent of programs

Adult Basic Learning Examination (ABLE): Used by nearly 10 percent of programs

Respondents reported using the assessment information generated by these tests for purposes including:

- To appraise basic skills in order to sort youth and assign them to appropriate programs (35 percent of programs);

- To diagnose where learning should begin within a defined level (70 percent of programs);

- To monitor progress (31 percent of programs); and

- To certify attainment or gain through use as a post-test (66 percent of programs).

One can infer that the most common assessment practice is the use of standardized tests for pre- and post-data collection. The next most widely reported assessment strategy was the intake interview, cited by 45 percent of the respondents.

When asked about issues or problems in implementing effective basic skills remediation programs under JTPA, all practitioners without exception digressed from the interview protocol to indicate that they regarded the lack of staff training in assessment and instruction as a serious problem. The next most often mentioned problems included "motivation and lack of incentives for participants", "attendance and retention," and "lack of cooperation from the school system." These three problems are also regularly raised by participants attending The Center for Remediation Design's Institutes on Basic Skills.

RELATED ISSUES

There are some developments beyond what we learned from the survey that are worth noting. Experimental programs are now underway in a number of SDAs around the country to determine the viability of using criterion-referenced tests rather than standardized tests. Criterion-referenced tests result in scores that indicate what the test subject can do, as compared to standardized tests, which result in scores that compare the test subject's performance to a representative group. If practitioner interest is any indicator, the trend may indeed be away from standardized testing and toward sophisticated criterion-referenced testing as a measure of the nation's employability.

For example, the major contemporary assessments of the basic skills of adults conducted by the National Assessment of Educational Progress (NAEP) describe what people know and can do; the NAEP assessments are intended to stimulate debate about whether those levels of performance are satisfactory. In the NAEP report (1987), the proficiency levels chosen for describing results on a proficiency scale ranging from 0-500 are: 150 - rudimentary, 200 - basic, 250 - intermediate, 300 - adept, and 350 - advanced. Each level is defined by describing the types of reading material and tasks that most "students" attaining that proficiency level would be able to perform successfully; each is exemplified by typical benchmark exercises. (See Figure 1) In the scale-anchoring process NAEP selects sets of items that are good discriminators between basic skill proficiency levels and that related to survival or employment, i.e. that are meaningful to client groups such as those served by JTPA. (The Comprehensive Adult Student Assessment System [CASAS], which is described in Appendix A, utilizes a similar proficiency scale for both reading and listening comprehension tasks of 150-150 but the same principles apply.)

FIGURE I

Levels of Proficiency 1/3

Rudimentary (150)

Readers who have acquired rudimentary reading skills and strategies can follow brief written directions. They can also select words, phrases, or sentences to describe a simple picture and can interpret simple written clues to identify a common object. *Performance at this level suggests the ability to carry out simple, discrete reading tasks.*

Basic (200)

Readers who have learned basic comprehension skills and strategies can locate and identify facts from simple informational paragraphs, stories, and news articles. In addition, they can combine ideas and make inferences based on short, uncomplicated passages. *Performance at this level suggests the ability to understand specific or sequentially related information.*

Intermediate (250)

Readers with the ability to use intermediate skills and strategies can search for, locate, and organize the information they find in relatively lengthy passages and can recognize paraphrases of what they have read. They can also make inferences and reach generalizations about main ideas and author's purpose from passages dealing with literature, science, and social studies. *Performance at this level suggests the ability to search for specific information, interrelate ideas, and make generalizations.*

Adept (300)

Readers with adept reading comprehension skills and strategies can understand complicated literary and informational passages, including material about topics they study at school. They can also analyze and integrate less familiar material and provide reactions to and explanaltions of the text as a whole. *Performance at this level suggests the ability to find, understand, summarize, and explain relaltively complicated information.*

Advanced (350)

Readers who use advanced reading skills and strategies can extend and restructure the ideas presented in specialized and complex texts. Examples include scientific materials, literary essays, historical documents, and materials similar to those found in professional and technical working environments. They are also able to understand the links between ideas even when those links are not explicitly stated and to make appropriate generalizations even when the texts lack clear introductions or explanations. *Performance at this level suggests the ability to synthesize and learn from specialized reading materials.*

naep

CHAPTER III

USING TESTING TO DEFINE YOUTH INTERVENTION NEEDS

Program planners and operators must know the extent to which a young person lacks basic skills to make decisions on whether to place them in a remediation component. This chapter presents:

- An employability continuum that can help program planners design intervention strategies that meet the needs of the JTPA youth population;
- The four key steps in an assessment process: appraisal and screening, individual diagnostics, monitoring and benchmarking, and certification testing; and
- Guidance on implementing an assessment process which provides information on whether a youth has any deficiencies at all, how deficient they are, and specifically what they do not know. This information becomes the basis for developing a training/remediation plan which increases skill levels.

AN EMPLOYABILITY CONTINUUM

The starting point for developing a good assessment strategy for youth must be the young people themselves, rather than a review of the literature on test alternatives. The question is not "What test to use?" but rather, "What assessment process best meets the needs of the target youth population and will provide information that assists in designing a participant's training plan?"

JTPA-eligible youth can be viewed on a continuum, beginning with those needing a substantial amount of training because they have multiple and serious barriers to employment and continuing through those who need some training but who have the fewest barriers to employment. The following continuum, which first appeared in the National Governors' Association's **Assessing Employability for Results**, (Curnan, Fiala, Lerche, 1985) is a useful representation of the range of JTPA-eligible young people:

EMPLOYABILITY CONTINUUM

PRE-EMPLOYABLE	NEARLY EMPLOYABLE	EMPLOYABLE

- Pre-employable youth are those who are most at risk of being chronically unemployed and who will require the most intensive set of services from the community. Youth appraised as pre-employable will test at less than a seventh grade level in math and reading skills (or between 199 and 214 on the point scale developed by CASAS). Their need for (or lack of) work skills can be assessed through interview questions (e.g., have you ever worked?) or through a short work-based activity.

  In functional terms (i.e., what can they do with what employment-related basic skills they have attained), a youth appraised at below seventh grade level may be able to perform tasks such as identifying amounts of money, printing legibly in ink, and recording date and time.

- Nearly employable youth report some work history and/or demonstrate some competency in pre-employment skills. Their basic skills capacity in reading and math will be appraised somewhere between the seventh grade and below ninth grade (or between 215 and 224 on the CASAS point scale).

  At this level a youth may be able to read and interpret basic measurement and numerical readings on measurement instruments, read and interpret instructions for safe use of equipment materials and machines, and fill out forms.

- Employable youth will be appraised at functioning at or above ninth grade in reading and math skills (or at or above 225 on the CASAS system). These youth will demonstrate some knowledge of occupational choices, the capacity to get a job, and some history of keeping a job. In functional terms, this means a youth may be able to recognize and interpret ratio and proportion; calculate with units of time; read and interpret written sequential directions in textbooks, manuals, and handouts; and write memos and letters.

Participants enrolling in the JTPA system enter at any point on the employability continuum; if all goes well, they move along until they are part of the employable group. Clearly, participants may enter and exit at any point along the continuum, and some do exit before they are fully employable.

Learning as they go, JTPA administrators and planners are having to make difficult decisions about program design, assessment, and curriculum when it comes to basic skills training. The demand for basic skills is well understood, but few SDAs feel they have the tools to make good programs happen.

The employability continuum presented here represents a starting point for planning and designing basic skills programs around the characteristics of the diverse group of participants. The most critical point, whether serving in-school or out-of-school youth, is to determine the skill levels of the youth in order to determine the intensity of training required.

The three-level approach represented by the employability continuum uses seventh and ninth grade levels as benchmarks for varying program design. This is based in part on lessons from educational theory, in part on CASAS field testing with employment and training programs, and in part on practical lessons from experience within employment and training. It is important to note that it is always hard to draw the lines at specific levels and it is, in the absence of definitive research on the subject, always somewhat arbitrary. We could, for example, make a strong case for using fifth and eight grade benchmarks as do literacy training programs.

According to the Basic Skills monograph prepared by Brandeis University, Center for Human Resources (publication pending, U.S. Department of Labor, Spring, 1988):

> When discussing reading -- and by extension the other basic skills -- many educators commonly divide the population into three groups: those with skills below the fourth grade level, those reading at a fifth through seventh grade level, and individuals who can read at the eighth grade level or above. While educators and employment practitioners are increasingly dissatisfied with grade level as a measure of ability, those common benchmarks can help practitioners divide the youth population into segments that reflect the need for different types of program designs. In general, the fourth grade reading level marks the transition from the process of "learning to read" to one of "reading to learn." Below the fourth grade level, students lack the basic decoding skills needed to read printed materials; above that point they are able to work more independently and can read well enough to locate information, combine ideas, and make inferences from relatively simple materials. A similar shift occurs around the 8th grade level, as students are able to deal with longer and more difficult materials. An eighth grade reading level is often considered the minimum standard for functional literacy, though again, there is some disagreement about what skills are "functional" in today's hi-tech society. On a more pragmatic level, an eighth grade reading level is also the common dividing line between young people ready to pursue their GED or enter skills training and those who need additional, preparatory basic skills instruction.

> Within the general youth population, the vast majority of young people fall within the upper two groups, though there are still significant numbers of non-readers. (Sticht, Functional-Context Education)

For JTPA practitioners, the high percentages of eligible youth who read below the eighth or ninth grade level carry two sets of meaning. The first and most common is the indication of the pervasiveness and magnitude of the basic skills problems among youth. As more and more studies have demonstrated, significant proportions of the population -- particularly those segments served by JTPA -- have difficulty performing the basic reading, writing, and computational tasks needed to compete in the labor market. However, the figures also highlight a second point: the diversity of basic skill needs among young people and the importance of recognizing that diversity in planning and designing basic skills programs. Hence the three-level employability continuum.

Second, JTPA administrators and planners, realizing they had a wide range of skills in the youth programs and only a single program design, constructed the three-level program approach represented by the employability continuum in order to afford the level of intensity required to reach those most in need.

THE FOUR ASSESSMENT STEPS

Assessment is a multi-step process; it produces information on a participant which can be used for many program purposes. Each assessment step provides information on where a youth "fits" on the employability continuum and how far that youth has progressed toward employability.

Testing is but one part of an overall assessment strategy. However, the appropriate use of tests is an invaluable technique that can contribute to effective information at each of the four steps of the assessment process.

The four assessment steps are:

1. Appraisal (Screening). This first step in the assessment process provides an immediate snapshot of an individual's current abilities. Although the information produced at this assessment level may lack specificity, an initial appraisal which identifies a youth's functioning level of basic skills and work skills assists program operators in deciding whether the next level of assessment -- diagnostics -- is necessary or whether the youth is ready for a set of program services which does not include basic skills remediation.

2. Individual Diagnostics. The more extensive assessment carried out in this step provides information on

specific skills in which a youth is deficient. This information pinpoints exactly where the remediation process should begin. This step functions as the "pre-test." At this point a fairly prescriptive employability development plan should be formulated.

3. Monitoring Progress (or Benchmarking). This assessment step provides program operators with information on how well a youth is progressing in the program and indicates when specific goals are met for the purpose of program exit. For the participants, benchmarking progress serves to reinforce learning by focusing on accomplished goals and specific competencies mastered.

4. Certification Test. This test is designed to verify competency attainment. This step functions as the "post-test."

USING ASSESSMENT STRATEGIES TO DEFINE TARGET GROUPS

In order to develop a basic skills remediation program, it is necessary to define the target group according to basic skills needs and work skill needs, rather than to rely on the more common approach of defining the target group based on demographics (i.e., offender, teen parent, ethnic group, etc.). This helps program staff make more effective decisions on who needs remediation. While those who do not need remediation may receive other JTPA services, this approach to targeting avoids inaccurate assumptions based on demographic characteristics (e.g., all teen parents need remediation simply because they are teen parents). This approach requires that some method be developed, even in the appraisal step of assessment, to make an initial determination of achievement levels.

Defining the target group according to basic skill deficiencies as related to occupational needs also enables planners to place individual youth on the employability continuum and to design cost-effective training strategies based on actual need rather than assumed need. Effective assessment strategies are practical and provide immediate information that can be used to develop each participant's service plan.

Once the appraisal has been completed and the youth has been placed on the continuum of employability based on basic skill and work skill deficiencies, a set of services can be identified for that youth depending upon where he or she falls in the continuum. In other words, specific services can be matched to a youth's needs which will assist in upgrading (or remediating) the skills the young person lacks. The schematic on the following page identifies the type of service needed at each level.

JTPA ELIGIBLE PARTICIPANTS

PRE-EMPLOYABLE

- Basic skill level: 7th grade or below
- General services needed:
  - Basic skill remediation
  - Work experience
  - Pre-employment skills development
  - and others

NEARLY EMPLOYABLE

- Basic skill level: Below 9th grade down to 7th grade
- General services needed:
  - Basic skill remediation
  - Work experience
  - Pre-employment/work maturity skills development
  - and others

EMPLOYABLE

- Basic skill level: 9th grade or above
- General services needed:
  - limited basic skills remediation
  - Job search assistance
  - Job specific skills training
  - and others

(These services are not listed in any particular order for delivery. They will be delivered concurrently or sequentially dependent on the individuals service plan.)

CHAPTER IV

OVERVIEW OF THE APPRAISAL PROCESS: TEST AND DATA GATHERING ISSUES

Testing is a useful technique at each step in the assessment process: appraisal, individual diagnostics, monitoring progress, and certification. The authors discuss:

- The uses and limitations of tests during the appraisal step of the assessment process;
- Data that can be used to supplement test results; and
- Testing issues regarding each of the five basic skills: reading, written communication, verbal communication, math computation, and problem-solving.

APPRAISAL PROCESS

Tests can only be effective to the extent that their users specify the purpose for which they are being used and understand their limitations. Equally important, test scores without additional information are not useful for most purposes. Consequently, well-designed assessment procedures typically integrate both non-test and test-based information.

The appraisal step of the JTPA assessment process is the first point at which testing is likely to be used. This step is designed to gather the broadest amount of relevant information in the most efficient manner in order to identify those clients who are at high risk of having specific basic skills deficits as well as those clients who probably do not have such difficulties.

(It should be noted that the word probably is critical in this discussion. All tests result in probabilistic statements; they do not provide hard facts. A common misunderstanding about the assessment process, and about test scores in particular, is that some unchanging and "true" measure of a clients abilities results. In fact, what this step provides is an estimate of a client's abilities, and there is always error in that estimate, mainly because there is no test which perfectly predicts any general skill, behavior, or ability.)

The use of an appraisal test identifies the "employable" clients for enrollment directly in job training activities such as job search, specific skills training, etc.; at the same time, it identifies those clients for whom additional assessment is

necessary in order to further define their basic skills deficits and to plan remediation. This is the most cost- effective way to identify those clients who need the individual diagnostics step without providing that step for all clients.

The initial appraisal step should include a structured interview which provides information regarding the clients' medical, educational, and work history. In addition, the interview should provide some information regarding the client's adaptive functioning and psychological/emotional state. Rating scales relevant for judging a client's presentation, verbal communication skills, and social abilities, can be very useful in the hands of trained interviewers and raters. Any relevant records from the client's school or work setting should be gathered before, or at the time of this interview, to document recent functioning.

If schools can provide test results or other information for JTPA clients, additional assessment data may or may not be needed for classification purposes. The type and quality of the data from the schools is of primary importance. All the issues regarding test content, reliability, and validity should be considered when evaluating the types of data gathered from a school.

Better data may be available from the schools than from standardized group achievement test scores. Actual work samples may be available in the areas required (written communication, mathematics, problem-solving) which could be rated using a systematic and reliable rating system. Ratings of basic skills from the client's teachers may also be available.

In general, psychological test results from school records which are more than two years out of date are probably not useful for current assessment purposes as additional learning probably has taken place, although they can document previous functioning and any changes over time, which may be useful for predicting success in various programs. It is probably most important that the appraisal interview collect similar information across all clients in order to provide useful information for individual client predictions and/or program evaluation purposes. Many times, interview data can be as predictive of skill deficits, or program success, as can standardized psychological test scores if the data is collected in a systematic and consistent manner, and then used in the development of screening and grouping criteria. Interviewers should be trained in a standard manner in order to obtain the most accurate and valid information from clients in the JTPA program. Interviewer skills are a key to the success of this component of the initial assessment process.

Overall, data from schools, combined with a comprehensive interview, may provide a useful alternative to an independent screening assessment within the JTPA program. Whether a program uses school derived data, its own assessment data, or some combination, the usefulness and accuracy of the data can only be derived through an evaluation of the program's results.

Although a comprehensive interview can provide much useful information regarding a client's history and current functioning level, it is sometimes very difficult for such an interview to measure basic educational/academic skills in a reliable and valid manner. Because of this, a general screening of such skills using standardized tests can be beneficial. However, there are several hundred psychological tests which purport to measure achievement levels and basic academic skills. Probably the most confusing aspect of test selection is that test names may not represent what they actually evaluate. For example, many "math" tests use word problems. Although word problems have traditionally been used to assess mathematics abilities, and probably relate to real life problems involving mathematical skills, a client who cannot read may score very poorly on such tests despite having adequate math skills. A test's name does not necessarily represent the abilities that the test assesses.

JTPA practitioners must make clear decisions regarding the purpose of such tests at this first step of the assessment process in order to select the most useful testing instruments. Numerous issues in addition to the test's psychometric properties must be considered. (Psychometric characteristics are discussed below.) These include:

- Testing time;
- Administrator qualifications;
- Test costs;
- Scoring difficulty; and
- Relevance.

The most important factor for JTPA purposes may be the question of occupational relevance, which is discussed later in this chapter in the context of testing for individual basic skills, and again in the final section of Chapter V.

Two aspects of the job placement process may require two different skill levels, and the appraisal step should begin to identify skill deficits regarding both these aspects:

- Obtaining a job and meeting its entry-level criteria may require those basic reading and writing skills necessary to apply for the job and to perform basic job training or entry level activities; and

- Retaining the job and progressing in it may require more different or more advanced skills than those required at entry.

The purpose of the appraisal step should be to identify those clients who lack the general academic skills necessary to obtain a wide variety of jobs. "Job specific" basic or more advanced skills in these areas should be assessed at a later point in the process. In order to screen a client's basic academic skills it is necessary to identify which types of skills are considered

basic and generic across most occupational situations. Once such skills are identified, then test selection becomes easier.

The Dictionary of Occupational Titles (DOT) may prove quite useful for practitioners trying to identify the general academic skills necessary to obtain a wide variety of jobs. The DOT focuses on occupational classifications and definitions by standardizing and defining job duties and related information for over 20,000 occupations.

The DOT classifies jobs into job categories, divisions within each category, and specific job titles within categories. Each classification level identifies the skills, knowledge, and abilities a person needs for the job. While the DOT is primarily designed as a job placement tool to facilitate matching job requirements and worker skills, the identification of worker functions is ready-made to help JTPA practitioners tie basic skills to functional skills. The definitions delineate how well a worker has to read, write, etc., by describing the way each basic skill is used to perform job functions. The DOT does not identify at what grade level a person must function (other than to specify certain certificates). The DOT focuses on the worker functions necessary to perform the job. A solid understanding of the relationship between worker functions and basic skills (assisted by the DOT) can assist practitioners in deciding whether a participant needs further diagnostic assessment and remediation.

(It is important to point out that tests can only sample the behaviors or skills which are being assessed. Their purpose is to predict a client's actual abilities in the real world. Whether a JTPA client can correctly answer 18 out of 20 math computation questions on a test is less relevant than whether that client's performance on the test corresponds to the computational abilities required on the job. Without such correspondence, a test serves no useful purpose.)

TESTING ISSUES REGARDING THE FIVE BASIC SKILLS

There are five basic skills which are considered to be transferable and important across most occupational areas. These include:

- Reading comprehension;
- Written communication;
- Verbal communication;
- Math computation; and
- Problem-solving.

The first three skills are all linked to basic language skills and represent a client's ability to understand written language, produce written language, and produce spoken language. The second two skills are also linked in that both include problem-solving skills of a conceptual nature.

Although this description of the five basic skills sounds very simple, assessing an individual's mastery of them is very complex. Each of these basic skills is made up of a variety of subskills, which may or may not be important in specific JTPA testing situations. Many times a test does not clearly identify which subskills it is assessing, and may give the impression that it is assessing all relevant subskills, although there are no such comprehensive tests available. (A good example of this problem is a "reading" test [i.e., WRAT-R] which only measures the ability to read words but not the ability to comprehend them.)

The following description of the five basic skills offers JTPA decision-makers a summary of each skill as well as an introduction to some of the subskills of each. This should help practitioners ask better questions when trying to select appropriate tests for a given purpose.

1. Reading Comprehension

Reading is a very complex ability with many different forms and subskills. Many so-called reading tests assess a client's ability to read single words/non-words (i.e., "Reading Vocabulary"). Such tests are typically described as assessing phonetic decoding skills (a reading subskill), or reading through sight (whole word) reading strategies. A client may have excellent decoding or single word reading skills and perform very well on such tests but have no comprehension (understanding) of what he or she has read. Other reading tests assess a client's ability to read sentences, paragraphs, or contextual information (i.e., "Reading Comprehension" subtests). Again, there are clients who may be able to read such texts without comprehension. Thus, they may get a high score despite lacking functional reading ability.

Various reading tests may measure very different reading subskills. Examination of actual reading tests shows that some require silent reading, while others require oral reading; some pose questions to determine comprehension levels, while others do not. Word type (phonetically regular or irregular), sentence structure (syntactically complex, etc.), or paragraphs (inferential, concrete, etc.) may also be different, and the level of vocabulary involved may vary. The type of response required in different reading tests may require pointing to a picture, retelling the content, answering questions, filling in a missing word, or writing an answer. Because of all of this variation, a client may achieve a high score on one reading test and a low one on another, and both may be accurate indices of his abilities.

Many relevant, related abilities and a great deal of knowledge also impacts reading skills. A client who has a very limited vocabulary, for example, typically cannot comprehend text which

includes words above his or her vocabulary level despite having the ability to decode ("read") the text. In this case, it may be misleading to interpret a low score on a reading comprehension test as being due to an inability to read rather than to limited language and vocabulary development. In general, any limitation in language development, or any of its subskills, affects reading abilities. Therefore, some language assessment is required in order to obtain a good diagnostic picture of any client with reading problems.

In general, reading tests used for appraisal purposes should assess reading comprehension. Ideally, the test should require the client to read (silently) paragraphs of increasing complexity; it should time the client's rate of reading; and it should pose questions about the content of the text for oral response. Such a test would be more "real life" than many of the other types of reading tests available, and would provide for a more global assessment of reading abilities than many other options. (This screening recommendation may not be appropriate for jobs requiring oral reading such as phone operator or dispatcher.) Various subskill deficits could cause a low score on such a test (poor single word decoding, weak vocabulary, poor memory, limited reading comprehension skills, poor attention, etc.), although the exact subskill/ability deficit resulting in the reading comprehension deficit cannot typically be discerned by such screening measures.

Identifying the cause of the reading problem is the goal and purpose of the individual diagnostics step of the assessment process. It is at this diagnostic level where designing the most appropriate remediation would also occur. What is most important in this regard is that almost any reading subskill relevant to "real life/job situation" reading ability, if deficient, could affect a client's score on such a screening test; further testing would be needed to identify such problem areas.

2. Math Computation

Mathematical and computational subskills and tests, like those related to reading, are numerous and typically multifactorial. Basic mathematical abilities include addition, subtraction, multiplication and division. Fractions, percentiles, decimals, money, time, and other types of measurements, are also included. Mathematical assessment also requires dealing with the issue of single-step or multiple-step problems, single-digit or multiple-digit problems, and mixed procedure problems (e.g., adding and dividing within the same problem).

As with reading, there are many different ways to test math skills; some tests require the client to compute and write out the solution to the problem, others use a multiple choice format, while still others require the solution of word problems (which assess a client's ability to understand the problem and solve it in addition to computing the answer). Some tests also require

computations within time limits. The assessment of geometry, algebra, trigonometry, and other mathematical areas is not generally useful in basic skills evaluation unless those topics are job-relevant (e.g., geometry in drafting). Finally, there are numerous mathematical concepts related to measurement constructs which may also require assessment if deemed job-relevant (yards, metric measures, quarts, etc.).

For screening purposes, it is probably most relevant to assess a client's skills in performing increasingly complex written computations, without time limits, for all major areas (addition, subtraction, multiplication, division, fractions, percentiles, decimals, money, time, and other measurements with single and multiple-step problems using multiple digit numbers). Such a general test would provide a "general mathematics" score, but would also provide some initial diagnostic information about a client's specific deficits.

3. Written Communication

Writing is one of the most complex of the basic skills. Written communication ability typically suggests that a client has adequate speaking and reading abilities since writing is based on initial mastery of those skills.

Because written communication skills are multidimensional, so is their assessment. Two traditional testing approaches have been used to assess written communication skills: 1) having a client write within prescribed guidelines (spelling, capitalization, punctuation tests), or 2) having a client produce a spontaneous writing sample (write a story on a specified topic). A client who performs well on prescribed writing tests may not have adequate ability to write in a meaningful and communicative manner, and vice versa. Unfortunately, most standardized achievement tests only assess spelling -- a limited subskill of written communication -- and cannot assess spontaneous writing skills except in an overly structured manner.

The following written language subskills should be considered when screening writing abilities:

a) Mechanical penmanship or handwriting skills (mechanical formation of letters, words, etc. and general neatness such as spacing, alignment, etc.);

b) Written language rule use (punctuation, capitalization, etc.);

c) Spelling;

d) Vocabulary, linguistic structures (syntax, grammar, semantic structures, verb tenses, plurals, subject-verb agreement, etc.); and

e) Logic of content and theme.

All of these subcomponents of written communication are interrelated although there are few standardized tests with norms which assess all of them in such an integrated framework.

An adequate screening assessment for JTPA programs should probably begin with an assessment of the more basic subskills in written communication (handwriting, written language rules, and spelling), while diagnostic assessments should focus on the more complex components of the written language act (vocabulary, linguistic structures, logic, and themes). Standardized assessment of these primary subskills combined with a spontaneous writing sample should be sufficient for initial screening purposes.

4. Verbal Communication

Verbal communication, or spoken language abilities, are closely related to written communication skills and subskills. While there are almost no standardized paper and pencil achievement tests which assess spoken language skills, there are various rating scales which can be completed by anyone who talks and interacts with a client. These ratings, which could easily be carried out based on the verbal behavior of a client during the initial interview, typically assess articulation skills (ability to speak clearly and intelligibly), level of receptive and expressive vocabulary, ability to comprehend another person's questions and statements (receptive comprehension skills), expressive fluency (amount and rate of speech), and appropriate use of linguistic structures (grammar, syntax, plurals, etc.).

5. Problem-Solving

The concept of "problem-solving" is probably the most difficult of the five basic skills to describe and define. There is no such thing as a single test of "problem-solving" abilities which covers all skills which most people consider under this topic. Typically, the idea of problem-solving addresses a group of interrelated skills which are utilized to deal with any new, complex, or abstract concept or situation. It includes the subskills of:

- Planning and organization,
- Goal setting,
- Appropriate use of feedback,
- Reasoning,
- Set switching,
- Information coordination, and
- Concept learning.

Some people suggest that there are two components in all problem-solving activities: understanding the problem, and being able to

solve it. Each of these subskills within the area of problem-solving are difficult to define and/or assess. These skills also overlap greatly with those involved in the other four skill areas. Some of these problem-solving abilities are involved in mathematical operations, as well as in much reading comprehension and written communication. In fact, this area could be considered a subskill to all other basic skills, because without it, the other skills are only automatized responses without generality and flexibility in new situations or problems.

In the screening situation, it would be best to identify problem-solving tests which neither depend on, nor assess, reading, writing, or mathematical skills. Preferred tests should assess reasoning and concept formation to be most useful in the appraisal process. Unfortunately, many of the standardized tests on the market which claim to assess problem-solving skills are only limited verbal analogies, or math word problem tests. It may be best to assess such problem-solving abilities in a work situation, or in more "real life" settings than via psychometric tests, as such skills are so complex and difficult to assess, and the scores in this area from psychometric tests are difficult to evaluate.

CHAPTER V

TEST SELECTION AND MEASUREMENT ISSUES

This chapter explores some of the psychometric issues that have a bearing on test selection. These are issues that program planners can use to balance considerations of cost, availability, or ease of administration that may otherwise limit the accuracy and utility of test results. The chapter covers:

- Advantages and disadvantages of defining deficiency relative to a population versus relative to deficits in an individual's own abilities;
- Test and measurement issues that affect the test selection process, including: validity, reliability, individual versus group testing, multifactorial tests, classification errors, normative data needs, use of grade-equivalent test scores, diagnosis, monitoring, and pre- and post-testing; and
- The relationship of testing to the delineation of job-specific skills.

DEFINING DEFICIENCY

There are two different types of deficiencies which test scores identify. The first is a deficiency relative to a population. The question addressed in this approach is whether the client is below a certain level on the test compared to the general population. A well-known example of this type of discrepancy is that involved in mental retardation on IQ tests. To fall in the "mentally deficient" range on the IQ test, a person has to score at or below a score of 69 (100 is average, and 69 or below represents the bottom 2 percentile of the population). There is nothing special about this score, and it has been decided arbitrarily. A score cut-off of 75, or 65 may be just as useful.

A deficiency definition relative to a population would probably be most useful in defining a JTPA client's deficiency. The biggest difficulty with using such a deficiency definition is in deciding on the most appropriate and useful cut-off score, and in deciding which population norms to compare such clients on for scoring purposes. Ideally, the scoring cut-off for determining a deficiency would be empirically derived through research by showing that clients below a certain level would be best served by one type of remediation program, and those above that level would be better served in another program. Such a criterion

could be derived over time at any JTPA site by adjusting the criterion based on feedback from the different programs and the success of the different types of clients. In many situations, the criteria for determining a deficiency are based on the level of special resources available to deal with the identified group. In other words, if a program had the resources to serve only 200 clients at a time in remedial reading comprehension classes, then the cut-off score could be set at a level which would identify the 200 clients with the lowest assessed reading levels as a percentage of the total number of clients assessed.

A major problem with the use of a specific cut-off score in defining an academic deficit is that traditionally a client whose general intellectual abilities were at a level consistent with his or her academic abilities would not be considered deficient in academic skills. In other words, a client ranked at the 15th percentile in intellectual abilities and at the 15th percentile in reading comprehension would not be considered deficient in reading. On the other hand, compared to the general population, that client would clearly be below the population average. One of the basic assumptions in this definition is that persons with low general intellectual abilities will not develop reading skills (or other academic abilities) at levels higher than their IQ, regardless of the remediation which may occur.

The other definition of deficiency is based on a relative deficit among an individual's own abilities. In this scenario, a client's abilities across all areas are compared, and any abilities which fall below the others are considered to be deficient. In this way, if a client's abilities in general fall at the 80th percentile level, but they show only a 50th percentile level of ability in the math computation area, this area would be considered to be a deficiency, even though the client's score may be in the average range for the general population.

Both definitions of "deficiency" are based on the concept of test score relativity. A score is deficient only relative to some other score -- whether that is a population average, or the client's own ability average. Given that most JTPA client's abilities are probably below the population mean, and that employers are probably not as interested in an individual's relative deficits, a population deficiency criterion is probably most appropriate in the JTPA situation.

The real question in this process is how to define the level of deficiency that requires a client to receive additional attention (assessment and/or remediation). The employability continuum in Chapter III of this report defined three classification groups: pre-employable (those with seventh grade or below basic skills); nearly employable (those with eighth to ninth grade basic skills); and employable (those with basic skills at or above the eighth grade level). This classification system also includes information regarding previous work history. The grade levels which define the employability continuum represent criteria which

have been assigned primarily for JTPA purposes rather than for academic purposes. They are useful in that they are not dependent on the client's age, previous educational level, or other relevant background; rather, they are based on a level of basic skills that such clients should possess as they move toward employability.

Criterion- and competency-referenced tests may be more useful than grade-equivalent scores in defining deficiency as they are not referenced to age or grade groups, but rather require the client to pass a specific test at a certain level of proficiency. By using such a testing system, which is frequently highly linked to instructional/remediation programs, there is little emphasis on an individual's grade level or percentile of abilities. What is important is that the client obtains a certain mastery of the basic skills necessary to move toward employability. Individualized and computerized instructional programs frequently use such mastery testing to assess client progress in a systematic manner. Unfortunately, criterion-referenced tests have yet to be developed to such a sophisticated level to be widely used and validated.

Defining basic skills deficiencies for JTPA clients is not an objective process. As in all such situations where testing and assessment data is utilized in the making of such decisions, there is the need for systematic collection of data by which to assess such definitions' validity and accuracy. The JTPA system must make a commitment to collecting such data to fully implement a system that targets the needs of the individual and provides programs that allow for levels of instruction and certification.

FUNCTIONS AND LIMITATIONS OF TESTING/ASSESSMENT

Many test and measurement issues may limit the utility of assessment tests. Some of these are technical issues involved in test construction, while others are related to the use of individual or group tests, problems involved in testing special populations (e.g., JTPA clients), and the limitations of standardized normative data for making predictions in special populations.

No test is perfect; all tests have limitations. However, if JTPA program operators consider the issues raised in this section when they are exploring which tests to use, they will increase the likelihood that the tests they select will be as accurate as possible if used appropriately.

Validity

Validity is the most important test and measurement issue. No matter how well a test is developed, and no matter what the test developers say the test does, a test is not useful for any purpose without some information about its validity.

"The validity of a test concerns what the test measures and how well it does so" (Anastasi, 1982). As noted earlier, the name of a test does not necessarily indicate what it is testing at all; most test names are overly broad (e.g., "a reading test"). Also, there is no such thing as a valid or a invalid test. A test is only valid for a specific purpose. For example, a test may be valid in predicting a client's performance in a given job training program, but may not be valid in predicting performance on a job. Likewise, a test may be valid for predicting grade point average in high school, but not for predicting intelligence. Any test is valid or not valid only in relation to the purpose for which it is used.. This is why, for JTPA screening purposes, the same test may not be valid for youth and adults, as the purpose for which it is given differs for those two groups.

There are different types of validity. Content validity concerns whether the test systematically evaluates a representative sample of the client's behavior in a given area. Content validity is important in JTPA basic skills screening because it determines whether the test is assessing a broad range of knowledge and skills within a given area. For example, a math test which only assesses a client's ability to add and subtract would have limited content validity if one were trying to predict general arithmetic skills. In addition, a mathematics test which required the subject to read complex word problems may be testing reading skills more than math. In order to judge a test's content validity, the user should refer to the specific test's development and standardization manual and review the actual test specifications and topics covered. In addition, a review of the procedures which guarantee content validity within these areas should be found in the same manual. Actual results from using the test should also provide some information about content validity, as scores on the test should get higher with increasing grade level.

A criterion-referenced test bases its assessment on whether or not a client has mastered a particular kind of information or skill (e.g., multiplication tables, or primary punctuation skills) and consequently is especially sensitive to problems in content validity. In evaluating such tests it is very important to ask whether the test covers a broad and representative sample of the skills under evaluation, and also whether or not the test is independent from the effects of other skills (e.g., reading word problems in a math test).

A second type of validity, criterion-related validity, concerns a test's degree of accuracy in predicting a certain behavior, situation, or skill. (This is not to be confused with criterion-referenced tests.) For example, a client's reading test score may be evaluated against the client's supervisor's rating of on-the-job reading ability. A reading test with high criterion-related validity would be able to predict the supervisor's rating. This type of validity is especially relevant when trying to make a

diagnosis about an individual via a battery of tests. The JTPA screening problem is a classic example of this question: is a client deficient in basic skills? Through a short but valid screening test, a prediction is made regarding this question. How accurate that prediction is is based on the results of a more comprehensive evaluation, review of all records, etc., and is a measure of the test's criterion-related validity in relation to predicted status (deficient or not). Criterion-related validity must associate a test score with independent criteria (supervisor's rating, teacher's evaluation, more extensive testing results, etc.) which it is trying to describe or predict.

Reliability

Reliability indicates consistency. A reliable or consistent test is one which yields similar scores from an individual client from one day to the next, providing stable scores over time. Reliability does not imply that the scores obtained are "right," but only that the test is measuring similar things today, tomorrow, and in the future.

The concept of reliability is related to the test-taker's mood, the amount of noise in the testing room, and a wide variety of other factors which are irrelevant to the ability being assessed but which may affect the individual's score on a specific test. The more reliable a test, the less these factors will affect the scores obtained. Thus, tests which are highly reliable should be less affected than less stable tests by environment and mood. It would be difficult to believe that a client could score at the eighth grade level in reading comprehension one day while scoring at the 12th grade level the following week. If this occurred, one would have to doubt the usefulness of such a test, as such results would suggest that environmental or mood factors, rather than reading comprehension ability, were affecting the test scores, as a four-year jump in reading comprehension over a week's time would be nearly impossible.

How reliable must a test be in order to be useful? For JTPA purposes, the higher the reliability, the better. However, there are two most common measures of reliability:

- Test-retest is an actual value showing how similar the scores are for a client who takes the same test twice during a specified time period. Typically, this time period ranges from a few weeks to a few months, although it can be two testing periods over a year or more. The longer the period between retesting, the lower the typical reliability "coefficient." Most adequate to good tests would have test-retest reliability coefficients between the values of .70 and .85; an excellent test would score over .90. As the coefficient nears 1.00, the reliability becomes closer and closer to perfect, suggesting little change in the relationships between scores over time.

- Split-half or alpha reliability is based on the concept that you can assess reliability within a single testing by comparing results across similar or different items within the same test. Although the actual mechanics of doing this are too technical for a discussion in this report, the scores obtained are similar to those described above.

Finally, regarding test construction and reliability, the more items or problems on a test, the higher its reliability. The main implication of this factor for screening tests is that, because screening tests are by design short and consist of few items, such tests are unable to result in any but the most preliminary findings. However, this is satisfactory because the purpose of the screening test is simply to determine whether or not the individual diagnostics step is needed in order to identify specific deficiencies.

Several testing limitations are based on reliability issues:

- First, the less reliable a test, the greater the error in measurement, which means there is greater error in the classification and grouping of clients;

- Second, the less reliable the test, the less valid it is, which means it doesn't measure what it purports to measure as well as it could; and

- Finally, reliability is not just a test-specific issue, but is relevant for any interview data or rating scales used to evaluate a client. Such data should also show consistency over time.

Overall, although reliability is a technical issue, it is a necessary consideration in the test selection process.

Individual versus Group Testing

Individualized tests and group tests serve different purposes. Most screening testing is performed in groups while diagnostic testing tends to be more individualized. The decision to administer individual or group tests should be based on whether there is a need to assess an individual client's ability or to describe a group of clients' abilities. Group testing is typically used in program evaluation and planning, while diagnostic testing is typically used for client specific remediation or placement. The reliability and validity issues discussed above are relevant to both types of testing.

In general, for JTPA purposes, the difference between individual and group testing has to do with trade-offs between cost and effectiveness in testing a large number of clients. Many of the major achievement tests available can be given either way.

Generally, group testing is considered a more efficient way to screen clients, but reliability and validity may be limited compared to individualized testing. Group testing also typically provides better normative data (discussed below) due to the number of clients which can be assessed at a given time. A major drawback in group testing is that it limits the types of items which can be asked (typically multiple choice, fill in the blank, etc.) and this may limit the nature and extent of basic skills which can be assessed in a group testing situation. In addition, group testing does not allow for an assessment of the individual's state while being tested, nor does it assist in identifying situations which may yield invalid results (marking the wrong answer sheet column, etc.) or provide direct behavioral observations of the client.

Given the increased use of computerized group testing, the recent availability of computer-administered group tests which adapt to the client's level of ability should be noted. These tests initially assess a client's ability level in a given area, and then adjust the difficulty of the questions to the client's initial level of performance. At this point, they perform a more global assessment of the client's abilities at his or her own level. Such computerized testing systems appear to be very "state-of-the-art" to those with limited knowledge in the testing area, but in fact they generally offer less development and standardization data than the more common paper-and-pencil tests. Studies have shown that users of these computerized tests report better and more "valid" results compared to paper-and-pencil test comparisons, even though the actual validity of the computerized testing systems may be much lower than the standard assessment strategies. Because many people are more impressed by a computer printout than by the quality of the data on it, it is important to assess any computer-administered testing systems carefully.

Multifactorial Tests

A client may score low on a given test for many reasons, because most basic skills tests and subtests are "multifactorial" in nature. This means that they measure more than one skill or ability at the same time, within the same scale. A math test which uses word problems is an example of such a test; in addition to measuring math skills, it also implicitly measures reading ability, computation skills, and problem-solving, all within the same test. A deficit in any one of these areas would result in low math score on such a test.

This issue is most relevant to devising individual remediation and training plans. In the example above, if a low math score resulted from the client's inability to read, reading remediation would be indicated, whereas if a low score was due to computation difficulties, mathematical remediation would be appropriate. Screening tests may not be able to separate such important diagnostic factors, but the individual diagnostics step of the assessment model would provide more information of this type.

(Multifactorial tests should not be confused with tests which have subtests (i.e., subskills) and form composite scores from various combinations of these subtests.)

Classification Errors Based on Test Data

Because tests can only estimate a client's abilities, there is the potential for error in making any decision based on test data. There are typically two kinds of decisions to be made during the appraisal step: that the client probably has a skills deficit, or he or she does not. There are also two types of decision errors which can be made: that a client placed in the "probable skills deficit group" is not skills-deficient (this is called a false positive error), or that a client placed in the "no skills deficit group" actually has a skills deficit (this is called a false negative error). On one hand, a client is placed in a situation without having the skills to succeed; on the other, the client is required to undergo additional evaluation to assess their deficits or to undergo additional remediation. Both situations have economic, personal and programmatic costs.

In order for the appraisal step to be effective, local SDAs must determine an acceptable level of classification error vis-a-vis the employability continuum. For example, while it may be acceptable that some clients may be mistakenly classified as having basic skill deficits, it is important that all clients who do have deficits be identified. This may be considered a liberal screening criteria, but it insures that all clients with any potential of having a basic skills deficit be identified. A simple example of this would be the use of a score at or below the 30th percentile on a reading test, rather than at or below the 10th percentile, as the criterion for a client to undergo the individual diagnostics step. Virtually all clients at or below the 10th percentile will have a reading skill deficit, while a few of those below the 30th may not.

Normative Data Needs, Test Bias, Special JTPA Population Needs

Normative data provide information for comparing and describing a client's abilities relative to some other group or criteria. A test score without normative data is completely useless. For example, saying that a client answered 20 items correctly on a mathematics test indicates nothing about whether the client did well or poorly on the test. The most common question asked about a client's score on a test is how that individual compares to others who have taken the test. Once a test is developed, it is administered to a standardization sample from which all norms are initially derived. Therefore, anyone who later takes this test can be compared to the normative group scores.

Interpretation of a client's score in relationship to this normative data curve can vary depending on how the test

developers wanted to describe clients' scores. Probably the most easily understood description is the percentile level, which represents the percentage of persons in the normative sample which the client scored at, or below. However, unless tests are developed using the same or similar normative samples, one cannot as easily compare across different test results using percentiles; thus, percentile scores may not always be as useful as they appear.

Other measures of an individual client's score which may occur are called standard scores, which typically have a mean of 100 and a standard deviation of 15. They can have any arbitrarily set mean and standard deviation. The use of similar standard score systems and normative samples makes it easier to compare across different tests and subjects. Almost all standardized achievement tests use standard scores and percentiles. It is interesting to note that on group achievement tests used in many school districts, a subject's score is given as a standard score and percentile compared to the national average (national normative sample), and as a standards score and percentile based on local school district norms. One's local norms may be higher than the national average, and a specific client's score may be lower than the local norm but above the national average, or vice versa. This example clearly shows the relativity of test score interpretation and the need to know the normative reference group to which a client's score is being compared.

There are several other ways (such as T-scores, CEEB scores, stanines, and deviation scores) in which test constructors measure an individual client's score against the normative group. While these methods can be complex, SDA decision-makers can generally develop a clear understanding of the scores obtained for a given test by referring to the test's administration and scoring guide.

The validity of a test, or set of tests, is very specific to the clients, situations, and purpose for which it is being used. If one changes any of these components, then the validity of the test may not be able to be generalized to the new group of clients, situations, or purposes. This is an important consideration in selecting tests for use in JTPA assessments, given that most achievement tests and basic skills tests were not developed specifically for the purposes for which JTPA may want to use them.

The use of any standardized test or assessment procedure is problematic if there is not an appropriate normative sample for comparison purposes. Such norms need to be for a group which is of similar age, socioeconomic, and possibly racial make-up. Comparing the typical JTPA clients to most national norms is probably inappropriate, given that those norms are usually representative of youth within the general population, but not of JTPA-eligible youth specifically. The question in such situations is whether such tests are valid for such a special population, and the answer is generally unknown.

In the best of all possible situations, all tests utilized with the JTPA population would be "re-normed" and validated on a sample from this client group. Without this process, all assessment and normative comparisons could be brought under criticism. The main concern for local programs is the amount of error which may occur in the assessment due to the differences between the normative group and the client group being assessed, and the fact that, consequently, the test may not predict, for a variety of reasons, the behaviors which it claims to predict. Because of this, the assessment results may contain a systematic bias, which, if it is not corrected, could cause clients to be classified erroneously and served inappropriately.

The best alternative to a restandardization and revalidation of assessment tests would be the keeping of systematic data on the use of such tests, their accuracy in classification and diagnosis in the JTPA program, and the establishment of clear guidelines for adjusting decision rules if biases or errors are found to occur. Keeping systematic data to further improve the assessment process would solve many of the problems being addressed in this report. Useful data elements would be:

- Age,
- Race,
- Sex,
- In school/out of school,
- Diagnosis match (i.e., did the diagnosis step confirm the screening results),
- Service received (i.e., remediation, not remediation),
- Client's progress/success rate,
- Number of years of school completed, and
- Diploma/equivalency cerfiticate/degree.

Disadvantages of Grade-Equivalent Test Scores

"Grade-equivalent" test scores have been commonly used in achievement tests (which is why this paper discusses deficiencies in grade level). Currently, most test developers are trying not to use such scores as they present difficulties in interpretation and meaning. Overall, grade level scores should be avoided if possible. They do provide a measure which is easy to understand but they are often misinterpreted and may not be accurate. The use of standard scores (discussed on the previous page) is preferred when available.

Grade-equivalent scores are based on the concept of a "typical" student's performance in a given grade. For example, a group of ninth graders is assessed on a test, their average score is figured, and this score becomes the score for an average ninth grader on this test. If a ninth-grade client takes this test and scores at this score level, that client is said to be functioning at a ninth-grade equivalency. If the client scores at a level below the average score, he or she may be described as scoring at

a "fifth-grade reading level." In fact, this is not the case. The ninth grade client may be scoring near the mean for the fifth graders, but this does not mean that that client's performance or knowledge is identical or similar to that of a fifth grader.

In addition, grade equivalents are not of equal "value" along the grade continuum. For example, a third grader who is performing at a first-grade level has a much greater deficiency than a eighth grader who is performing at a sixth-grade level, even though each student is performing two years behind grade level. This difference is due to the cumulative knowledge, and different abilities being tapped at different grade levels.

DIAGNOSIS

The purpose of diagnostic assessments is to provide individualized information regarding a specific client's strengths and weaknesses, and to provide information for remediation of any deficiencies. Although the data from the appraisal step can be useful in this regard, most appraisals are neither extensive nor specific enough to identify the causes of a client's deficits, or to provide details for a remediation strategy. Individualized testing/assessment may be required, although criterion-referenced teaching programs may provide a useful alternative to the need for extensive additional diagnostic work-ups.

The individual diagnostics step must be linked to the remedial training procedures being utilized. Most teaching strategies are based on a skill development model, and therefore require specific types of skill assessment as an aid in developing the remediation plan. Because of this, it is important for those developing individual diagnostics procedures for deficient clients to work closely with those involved in performing remediation with these clients. The diagnostic information obtained should be of high quality (reliable, valid, etc.), but also useful for the remediation program. Because there are so many different approaches to remediation, and because clients may have a wide variety of specific or global deficiencies, further guidelines, besides those already presented, cannot be specified for the diagnostic process. In general, this process does require a more trained and qualified diagnostician who is experienced with the clinical assessment of individual clients, and who is qualified in making interpretation of such assessment data. JTPA decision-makers will find Appendix A useful in working effectively with diagnosticians.

MONITORING PROGRESS

Unfortunately, many programs which gather screening and diagnostic data on their clients do not use these data for program monitoring and further program development. It should be emphasized that these data are well designed for this purpose.

In addition, re-testing participants at later points in the program can provide useful information regarding participants' progress which can be used in judging the program's effectiveness. Based on such data, modification in the program can be made.

PRE- AND POST-TESTING

The re-assessment of a client's progress, using the same tests has a number of problems. First, by giving the exact same tests, a client has already had one experience with the material and items presented and may remember some of them during the next testing, thus inflating their scores while not actually having improved their skills. Because of this, many test developers have created alternative forms of the same test. By using different forms (which typically contain different items and problems) one can be less concerned with the issue of retest score inflation.

Another problem with reassessment is that teachers may "teach to the test." Given that teachers may know what test is being used to evaluate their clients and their programs, they may emphasize remediation of those skills which the test assesses, thereby raising their client's scores although not improving their more general skills. Because of this problem, it may be necessary to use different tests at pre-entry assessment as compared to post-program assessment. This strategy can be problematic due to the problems in comparing results across different tests which may have different content, types of items, or normative foundations. One way to minimize this problem is to change the tests used every few years so that teachers are forced to focus on teaching general skills rather than teaching to a specific test.

DELINEATING JOB-SPECIFIC SKILLS

Assessment within occupational training programs has additional requirements compared to those of more generic basic skills evaluation. As has been described, there are those basic skills such as reading, writing, and arithmetic which are generic across almost all vocational areas, but there are also specific skills which may be job-related (oral reading for a radio announcer, special geometry for a drafter, etc.). Although most of the generic basic skills have been described above, job-specific skills and their assessment have not been addressed. However, the generic basic skills have been tied to "employability" through techniques such as selecting tests which measure real life functions like reading comprehension, defining the skills in a local labor market which cross many occupations, and selecting tests which measure these skills.

Almost every type of psychological test or assessment tool could be useful in specific occupational programs. Unfortunately, many times skills irrelevant to the job are also assessed and such

results may inappropriately influence job program/placement decisions. Such issues have become even more important in the recent past as invalid tests or ones which assess factors not relevant to the job may exclude minority or special groups from specific jobs. Such bias in job specific testing has come under greater scrutiny.

Conducting a job analysis is one of the most common procedures for identifying job-specific skills. The job analysis should identify the specific job requirements and subskills, and other abilities needed by workers in a specific job situation. The selection of appropriate tests (reliable, valid, etc.) to assess the various components of the job, described through the job analysis process, is then performed. Such tests may not be traditional paper-and-pencil methods, but may depend on a client's performance on specific job samples, or simulations. Again, the validity of the procedure (i.e., whether it measures what it claims to measure) is the key, not its apparent relevance to the job at hand.

Finally, it may be possible to produce a sub-classification of jobs which have interrelated skills and abilities. In this way, all mechanical jobs may require a specific subset of special skills, while fast-food workers may require assessment for a different set of specific skills. Through such a job classification schema, it may be possible to assess a client's generic basic skills, and then their specific subskills in a general occupational area (i.e., mechanics) without a lengthy and overwhelming assessment sequence.

The Dictionary of Occupational Titles (DOT) was briefly discussed in Chapter IV. Before conducting a local job analysis, it would be useful to identify by job title the specific jobs in the SDA which may be available to JTPA participants, and then to define those jobs using the DOT. Within very specific jobs -- identified as job titles in the DOT -- worker functions and some of the basic skills necessary to perform them are identified.

The previous discussion reviewed basic psychometric issues in test selection and use. These issues are independent of those involved in selecting assessment measures which evaluate certain abilities. Unfortunately, even though one may be able to find an assessment tool which reliably and validly assesses the JTPA clients, there are even more complex issues in deciding on how to classify such clients for further assessment and remediation.

Table 1, adapted from Standards for Educational and Psychological Tests, jointly published by the American Psychological Association, the American Educational Research Association, and the National Council on Measurement in Education, provides standards by which most tests are developed and validated and provides a useful summary of the information in this chapter. These standards provide guidelines for test selection and test use.

TABLE 1: STANDARDS FOR THE USE OF TESTS

1A. A test user should have general knowledge of measurement principles and the limitations of test interpretations.

1B. A test user should know and understand the literature relevant to the tests he uses and the testing problems with which he deals.

1C. One who has the responsibility for decisions about individuals or policies that are based on test results should have an understanding of psychological or educational measurement and of validation and other test research.

1D. Test users should seek to avoid bias in test selection, administration, and interpretation; they should try to avoid even the appearance of discriminatory practice.

1E. Institutional test users should establish procedures for periodic internal review of test use.

2A. The choice or development of tests, test batteries, or other assessment procedures should be based on clearly formulated goals and hypotheses.

2B. A test user should consider more than one variable for assessment and the assessment of any given variable by more than one method.

2C. In choosing an existing test, a test user should relate its history of research and development to his intended use of the instrument.

2D. In general a test user should try to choose or to develop an assessment technique in which "tester-effect" is minimized, or in which reliability of assessment across testers can be assured.

2E. Test scores used for selection or other administrative decisions about an individual may not be useful for individual or program evaluation and vice versa.

3A. A test user is expected to follow carefully the standardized procedures described in the manual for administering a test.

3B. The test administrator is responsible for establishing conditions, consistent with the principle of standardization, that enable each examinee to do his best.

3C. A test user is responsible for accuracy in scoring, checking, coding, or recording test results.

3D. If specific cutting scores are to be used as a basis for decisions, a test user should have a rationale, justification, or explanation of the cutting scores adopted.

3E. The test user shares with the test developer or distributor a responsibility for maintaining test security.

4A. A test score should be interpreted as an estimate of performance under a given set of circumstances. It should not be interpreted as some absolute characteristic of the examinee or as something permanent and generalizable to all other circumstances.

4B. Test scores should ordinarily be reported only to people who are qualified to interpret them. If scores are reported, they should be accompanied by explanations sufficient for the recipient to interpret them correctly.

4C. The test user should recognize that estimates of reliability do not indicate criterion-related validity.

4D. A test user should examine carefully the rationale and validity of computer-based interpretations of test scores.

4E. In norm-referenced interpretations, a test user should interpret an obtained score with reference to sets of norms appropriate for the individual tested and for the intended use.

4F. Any content-referenced interpretation should clearly indicate the domain to which one can generalize.

4G. The test user should consider alternative interpretations of a given score.

4H. The test user should be able to interpret test performance relative to other measures.

4I. A test user should develop procedures for systematically eliminating from data files any test-score information that has, because of the lapse of time, become obsolete.

From: Standards for Educational and Psychological Test, American Psychological Association, 1974.

CHAPTER VI

TYPES OF TESTS AND THEIR USES

Several types of tests have been referenced earlier in this paper. This section summarizes explains the different types of tests, their strengths and weaknesses, and some basic guidelines for their use within a comprehensive assessment system. The practitioner will learn:

- The various types of standardized tests -- intelligence tests, aptitude tests, achievement tests, personality tests, interests and values measures, and occupational skills tests -- and something about their long history of development, use, and success in a variety of settings;
- The relative advantages and drawbacks of criterion-referenced testing, their practical and theoretical links to JTPA issues, and how they are used in competency-based programming;
- Which types of tests best measure which types of aptitudes, skills, and characteristics;
- Basic principles of interpreting test results, and the limitations of such interpretations; and
- The advantages of using a battery of assessment procedures rather than a single test.

FORMALIZED/STANDARDIZED TESTS (PAPER AND PENCIL)

Many different types of tests fall within the category of standardized tests (see earlier discussions of standardized tests in Chapter II and IV), and many have a long history and are highly developed. These tests may be given in groups, or individually, depending on the behaviors being assessed, their scoring requirements, and the complexity of their interpretation. There are six major types of formalized/standardized test which are of the typical paper/pencil format. These include tests of:

- Intelligence,
- Aptitudes;
- Achievement;
- Personality;
- Interests; and
- Occupational skills.

Because standardized tests are so common in our society, almost all JTPA clients have had some experience taking them. Measures of intelligence, aptitudes, and achievement are the best developed and most widely utilized, while occupational measures are less well developed but widely used nonetheless. Personality and interest measures have more difficulties in their development, interpretations, and validation. The JTPA practitioner would be most likely to use achievement, aptitude, intelligence, and occupational measures, in that order. Personality measures could possibly screen for psychopathology, while interest inventories could be useful in identifying occupational interests in clients.

Computerized administration of any of these standardized measures is possible, and many have been computerized for presentation and scoring. Computerized testing has the advantage of ease of administration, scoring, and sometimes interpretation. The use of a computer does not improve the test or increase its validity in any way, although many users of such services erroneously report more valid results due to their awe of computers.

Intelligence Tests

Most standardized IQ tests require individualized administration, do not use paper/pencil formats, and require high qualifications in those making interpretations (e.g., WAIS-R, WISC-R, Stanford-Binet, etc.). There are, however, a few paper and pencil tests which provide more easily obtained IQ assessments (e.g., Raven's Progressive Matrices, Multidimensional Aptitude Battery). These tests can be useful in estimating a person's general level of intellectual functioning within the normal population. They provide the most broad prediction of a person's functioning across almost all areas.

IQ tests were originally designed to predict academic success in typical school environments. Since that time their use has been widely expanded, sometimes inappropriately, to predict various other abilities. Limitations of such scales are related to how they were validated, the number of sub-skills they actually assess, and what they were developed to predict. If results of such tests are interpreted by school or clinical psychologists, their scores may be useful in predicting overall cognitive and academic functioning in the context of JTPA programs. Given the highly emotional nature of the debate regarding racial and ethnic differences in IQ scores, and the known cultural biases within some of these tests, careful selection and use with JTPA clients would be warranted.

The interpretation of most IQ tests is based on some standard score metric, typically one using 100 as the normative sample mean, with a standard deviation of 15. Thus, a client who scores 100 on the test performs the test as well as 50 percent of the subjects in the original standardization sample for their age.

If someone scores 85, then they are performing at a level comparable to 16 percent of their age peers, while someone scoring at 115 is performing better than 84 percent of their peers. One would typically expect, if all conditions were optimal, that such clients should score at similar levels on achievement or other ability tests. A client who scores at the 16th percentile on a IQ test, and also at this level on reading and mathematics tests, probably would not be expected to progress significantly above this level, regardless of the remediation programs provided. It should be noted though that all abilities can be improved with training and remediation, so that no score level is considered "permanent," although the amount of change from a given level is probably restricted. In general, the concept of IQ is probably over-rated and over-interpreted in today's society. These tests should be considered as tapping a wide-range of abilities and providing an "average ability score" which may, or may not be a useful index of a client's general functioning across many different areas of his or her life.

Aptitude Tests

While IQ tests provide only a single, global score, aptitude tests yield information on clients' more specific and different abilities. Many aptitude tests are actually "test batteries," comprised of many different single ability tests, in which a client's performance across all tests (subtests) is interpreted so as to identify specific strengths and weaknesses relative to the standardization norms as well as to the client's own abilities. Recently, multiple aptitude batteries have been developed to better assess a wide variety of abilities for both occupational and academic purposes (e.g., Differential Aptitude Test, General Aptitude Test Battery). Of course, such batteries can only assess a limited number of different abilities, and there is little choice by the user of which combination of abilities are included. These multiple aptitude batteries are typically standardized on the same population which makes their subtest interpretation and comparisons more accurate.

Achievement Tests

Achievement tests are the most widely used of all types of tests. They are designed to assess the impact of educational programs on students. Achievement tests generally assess what a client has learned, as compared to aptitude tests which assess a client's potential for learning. Some achievement tests, however, can be used for both purposes.

There is a common error in defining achievement tests as assessing the effects of education, while aptitude tests somehow assess a client's "inborn ability or innate capacity" aside from education. In reality, both achievement and aptitude tests measure what a client has learned in the past, and only to the extent that past performance is predictive of future learning can aptitude tests be successful. In general, achievement tests

assess a client's current academic functioning, while aptitude tests predict future learning ability. However, as the same test may be valid for both purposes, such terminology differences often leads to confusion and misuse of test results.

The major advantage of standardized achievement tests is that they provide an objective and consistent measure of academic abilities. They can provide data on which to identify clients who are not learning within the current framework, and they may provide some information useful in remediation, or adapting instruction to the individual. Such testing can also provide feedback to clients about their progress in a given program, and may influence motivation.

Most standardized achievement tests are actually achievement test batteries, which encompass multiple academic skills. Most of the nationally used systems (Metropolitan Achievement Test, California Achievement Test, Iowa Achievement Test, etc.) assess word reading (decoding), reading comprehension, spelling, computational math, and some general or specific areas of knowledge such as science or social studies. Others include vocabulary, punctuation, mathematical concepts, mathematical problem-solving, and related skills. These tests are typically group administered, and their results are relatively easy to interpret. Both national and local norms may be provided for comparisons. The real limitation of such measures is their lack of clear and easily derived information related to appropriate remediation or teaching strategies for the client involved. There may also be limitations regarding the functional validity of such test results in job related activities.

Personality Tests

Probably the second most widely used standardized psychological tests, after achievement tests, are those which assess personality or emotional functioning. Most of these tests are paper and pencil in nature, although some of the most widely publicized are not (ink blot tests). These tests all attempt to assess a clients' emotions, motivations, attitudes, and social functioning, although most use very different theories about what personality is, what it includes, and how you measure it. Some of these scales assess only psychopathology and not normal personality factors. A few of these tests also assess test-taking attitudes, response styles and biases (faking good or bad on a test). As can be expected, these tests typically have significantly more problems in their development, their interpretation, and their validity. Almost all of these tests can be biased (faked) by a client who has no vested interest in the process, or its results.

For the JTPA practitioner, these tests may provide screening for severe psychopathology. Their usefulness in assessing a client's motivation, attitudes, etc. may be very limited. Such tests are easily administered and scored by a certified administrator such

as a clinical psychologist or psychometrist. Interpreting results from such tests, however, is generally complex and requires professional assistance.

Occupational Skills Tests

Many different approaches are currently used to assess occupational qualifications. The use of systematic job analyses, job samples, simulated work trials, or the use of the assessment center techniques have all had support and use. In addition, some occupations require the use of special aptitude tests, which may assess psychomotor, mechanical, clerical, or computer-related abilities. The validity of each of these procedures for a specific job situation is the key to its usefulness. Although tests that seem "job-like" may sometimes motivate clients, such similarity to actual work does not necessarily make the test valid in predicting success in a given job. A client's actual interest, knowledge, and motivation regarding a specific occupation is typically not included in such occupational "skills" testing although such factors may affect job satisfaction and ultimate job stability.

The difficulty with such testing procedures has to do more with the test's development than with its administration. Identifying relevant job-specific skills is difficult, and there are few widely used standardized tests (except for Government Service). Most occupational tests are developed for a specific occupational setting, or for a given population (such as JTPA clients). The more widely used specific aptitude tests are very similar to the ability or aptitude tests described previously, although most in the occupational area assess fine-motor (dexterity, tool use, typing, etc.) or related abilities. These tests typically are easy to administer and interpret, but their specificity makes them useful only for given activities within an area. However, some aptitude tests are designed to measure more general math and reading skills, although these test results are also functional in nature.

CRITERION-REFERENCED TESTING AND COMPETENCY-BASED PROGRAMS

Criterion-referenced tests indicate how well an individual performs relative to some criterion or specific learning objective, rather than how well he or she performs relative to others, as with standardized tests. Standardized tests provide scores that allow simple comparisons between individuals, schools, programs, districts, labor market areas, states and nations; they are easily administered and take little time away from instruction; and with a long history of use by assessment experts and institutions, they carry scientific credibility. Yet there is wide agreement among practitioners that standardized testing sanctifies trivial forms of knowledge, suffers from cultural bias, and -- at best -- often provides incomplete and misleading information having an adverse effect on curriculum and

instruction. The argument for criterion-referenced testing is not necessarily an argument against standardized tests as much as it is a call for good tests, for honesty in testing, and for a clear statement of what the results measure and what they mean in an employment context. In effect, criterion-referenced testing as a part of competency-based training offers employers an opportunity to create their own "Workskills Report Card." Figure 1 illustrates how several major components of competency-based programs incorporating criterion-referenced testing compare with conventional non-competency-based programs.

As Figure 1 illustrates, the implementation of a genuine competency-based program requires a commitment to integrated and systematic planning, implementation, and evaluation of the training processes. There is only one type of criterion-referenced testing in competency-based basic skills programming, although many practitioners, trainers, and educators confuse competency-based programming with minimum competency testing where the emphasis shifts from teaching to testing. Competency-based programming connects what is taught to what is tested through a curriculum management system. A bona fide competency-based program therefore has all the qualities identified in the competency-based column in Figure 1.

During the last 10-15 years there has been an increased interest in insuring that students graduating from our public schools can read, write, and do math at basic levels (basic skills competency). This led to the concept of the minimum competency test, which was designed to assess a student's basic skills. The individualized competency assessment grew from this movement. Almost all of these processes attempt to assess particular educational skills in real-life activities or in special academic settings (e.g., adult-education classes). At times, they have been used to assess various public school students' competencies also (seventh grade promotion examination, for example). The use of computerized adaptive testing (CAT) programs in conjunction with such procedures makes them very attractive to the user.

Probably the biggest strength of these procedures are their links to specific remediation or educational programs based on a students performance. The original concept behind this assessment process would be that a narrowly defined skill would be identified, a series of specific problems would be developed to assess whether a client had mastered the concept, and, if mastery had not been achieved, particular remediation approaches would then be used to assist the client in passing this component of the system. If the client passed, then he or she would move on to the next concept. The results of this type of assessment is very easy to interpret, in that the client either has or does not have the concept/skill. The problem with such systems is their need to provide an appropriate educational component to help the client master the concept or skill. There are few systems which have such a interlinked assessment/education program.

FIGURE I: PROGRAM COMPONENTS OF COMPETENCY-BASED AND NON-COMPETENCY-BASED PROGRAMS

Program Components	Competency-Based Programs	Non-Competency-Based Programs
1. Desired outcomes*	Specific, measurable statements; typically at an objective level	Non-specific, not necessarily measurable; typically goal-level statements
2. Instructional content	Outcome or competency-based	Subject-matter based
3. Amount of time provided for instruction	Continue until participant demonstrates mastery	Fixed units of time (e.g., semester, term)
4. Mode of instruction	Emphasis on instructor as facilitator of participant performance. Uses a variety of instructional techniques and groups	Emphasis on instructor presentation
5. Focus of instruction	What the participant needs to learn (especially related to employability and employment)	What instructor is able and likes to teach
6. Instructional materials	Several different texts and media based on the various learning styles of the participants in the program	Single sources of materials (Text and/or workbooks)
7. Feedback on performance*	Report results immediately after performance in understandable terms to the participant	Delayed feedback
8. Pace of instruction	Paced to each individual's rate of learning	Instructor or group paced
9. Testing*	Criterion (competency) referenced--test measures participants' progress toward attaining intended outcomes	Norm referenced--based on relative performance of others
10. Exit criteria*	Participant demonstrates the specified competencies	Final tests and grades

CALIFORNIA STATE DEPARTMENT OF EDUCATION, 1987

The CASAS Example

One example of such a program is the CASAS (Comprehensive Adult Student Assessment System) system. CASAS presents itself as a "comprehensive educational assessment and curriculum management system." It was designed specifically for adult or alternative educational groups (English as a second language, etc.), and was nationally validated. This system provides day-to-day instructional direction, more general curriculum planning, and on-going assessment linkage. This system has identified 34 competency areas and has also been linked to vocational skill competencies. (A further description of CASAS is found in Appendix A.)

Characteristics of Criterion-Referenced Tests

Competency-based educational approaches suggest that they are more education-driven than assessment-driven programs. These programs purport to teach those skills the client needs, and are focused on a total service framework within the context of the client's goals. The practical framework that supporters of competency-based educational systems describe consists of an integrated management, guidance and instructional staff, all working to help the client meet their goals.

Problems with such systems are sometimes related to: their costs, the generalizability of their results to other situations and environments, their reliability, their validity, and problems in choosing the most appropriate system. (However, there are not many competitors with high quality, proven systems at this time). Probably their greatest limitation is their breadth, which is at times limited to reading and mathematics related skills. As these are new programs, and have not had the extensive history of the more formalized measures described above, it is not surprising that they have a number of scientific and practical weaknesses. Their specific development, at times, for use with non-traditional students and adults makes such systems custom designed for JTPA programs and their clients.

In employment and training systems, we have found that traditional standardized assessment indicators by themselves communicate very little about the quality or substance of employment-related basic skills. We have come to realize that a useful, valid and reliable assessment system must be based on practical achievements considered by employers to be significant and meaningful.

Competency-based programming is not merely the use of one isolated competency-based concept without integrating it into a total service approach -- for example, criterion-referenced pre-tests without reference to participant goals or program components.

For employment and training programs, criterion-referenced testing makes sense because it is based directly on employers' needs and establishes in employment related terms, exactly what the participant can do. Vague references to "gain rates" per 100 hours leave unanswered the question: what can he or she do with a grade level score of 5.5? The problem of measuring basic education skills in or out of schools is compounded by the fact that there is no universally used measure of "basic skills;" that various measures of "basic skills" give different numbers; and the discrepancies are likely to be largest for the lower ends of the scales -- precisely where the employment and training practitioners are working.

With grade level gains as the measure -- even in the best of conditions -- the question remains: what does it mean in terms of actual new competence to say that a participant has a gain rate of 2.2 per 100 hours? The greatest advantage of criterion referenced testing in employment-related basic skills programs is that we know exactly what competencies have been achieved. Sticht recommends that we move in this direction and offers specific advice:

> "Because of the incomparability of grade level scores from test to test, the inadequate characteristics of reading level scales, and the lack of a proper understanding of what "grade levels" indicate about competence -- of either youth or adults -- major contemporary assessments of basic skills of adults, such as the National Assessment of Educational Progress (NAEP) and the Comprehensive Adult Student Assessment System (CASAS) are using the more powerful psychometric methods based on item response theory." (Kirsch and Jungeblut, 1987; Alamprese, 1987).

These new assessment methods are not only appropriate for basic skills remediation programs under JTPA but incorporating this "criterion-referenced" method of assessment will inject specific meaning into the program for participants, instructors, and employers alike. For the participants, employers, and program operators, the advantages of criterion referenced assessment inherent in competency based programs far surpass the drawbacks. For the participants, immediate feedback and recognition and reward for learning provides motivation to continue. Multiple options to apply and demonstrate mastery enables various learning styles to emerge. Program operators and instructors have detailed assessment data enabling them to target day-to-day relevant instruction and to certify attainment of competence based on specific mastery tests. Employers benefit through the specific definition of competency in labor market terms versus the vague or even trivial relationship of grade level or standardized tests scores.

The barriers to implementing genuine competency-based programming under JTPA seem to revolve around a single issue: an increased management burden. Without question two specific aspects of

management are affected: staffing and record keeping. There are insufficient staff in most programs. Ratios of staff instructors to participants is often as great as 1 to 50 or more. This pattern leaves little room for individualized programming with regular assessment. For effective competency-based programs, the ratio must be closed to 1 to 12 or less. Record keeping requirements do increase as the assessment process is ongoing, no longer an unrelated pre- and post-test function.

APPENDIX A

TEST DESCRIPTIONS

The Buros Mental Measurement Yearbook, Ninth Edition, lists 68 achievement tests, 97 reading tests, 46 mathematics tests, and 100 intelligence and ability tests. There are also over 350 personality and 295 vocational tests. This appendix presents information on some of the tests most commonly used in JTPA programs as well as some new tests which may be of assistance. Remember that just because a test is widely used does not mean that it is a good test!

The listing provides a publisher's name (some tests are sold through numerous publishers, so this is not a sole source), information on when the test was normed (how old is it), the age and grade levels for which it was intended, the kind of scores available from it (scaled scores, percentiles, etc.), how scoring is done, cost per client, and administration information. Most of these tests on this list can be administered in a group setting, but some require individualized assessments. Many tests now have computerized scoring systems which may give various amounts of data and interpretation. Be aware that many persons equate the number of computer pages with the amount of important information (and valid information) being provided.

Note that just because someone may be able to give a test and score it does not mean that that person can interpret it appropriately. In general, the more a test is used for diagnostic rather than screening or monitoring purposes, the more training is required for adequate interpretation of the results. In general, psychometricians, educational diagnosticians, school psychologists, educational psychologists, and clinical psychologists are trained in testing and interpretation issues. Each state regulates these professions to maintain appropriate educational and clinical training. Check the credentials of any professionals you plan to hire.

The titles of the various areas the test claims to measure (subtests) are also listed. In general, all of these tests provide scores for each subtest as well as some type of composite scores across an area (such as reading). Such subtest scores are usually less reliable and valid than the composite scores but may assist in interpretations. These listings represent the publishers' titles and do not necessarily identify the actual abilities being tested by such tests. A typical example of this are so-called "Problem-Solving" subtests, which do not really assess problem-solving abilities in the global sense, but only assess a client's ability to solve word math problems.

Most of the test descriptions have a "comments" section which provides a brief overview of the test, whether the norms would be appropriate for JTPA clients, and a recommendation of the test's best use (e.g., screening, monitoring, diagnostics). In addition, a rating suggests how well the test screens basic skills in each area (adequate, limited, very limited). This rating is related to the type of test (i.e., reading comprehension and word recognition assessment is required for a test to get an adequate rating in the reading area), and what it covers, not whether it measures its topic well or whether it is reliable or valid. No judgments are made regarding which are "good" or "not good" tests for JTPA purposes. These listing are not a professional recommendation and the information provided can be obtained from publishing company catalogs and reference materials. The selection of any test for any purpose is best made with professional consultation and assistance. There may be other tests which are appropriate to add to this list.

Test: IOWA Test of Basic Skills (1985 Edition)

Publisher: Riverside Publishing Company
8420 Bryn Mawr Ave.
Chicago, Ill. 60631

Norms: Re-normed in 1985 (Forms G & H). Grade K-9, ages 5-14. Special norms for large cities and low SES schools are available. Scores available: standard scores, grade equivalents, percentiles, normal curve equivalents (NCE).

Administration: Can be given by teacher or other trained persons. Testing Guide available. Basic battery requires approximately 135 minutes, complete battery 255 minutes. Hand or computerized scoring is possible.

Cost: Approximately $4.00 per student.

Subtest Areas: Vocabulary, reading comprehension, spelling, mathematics concepts, mathematics problem-solving, and mathematics computation. Areas within complete battery include capitalization, punctuation, usage and expression, visual materials, reference materials.

Comment: Norms are too young for most JTPA clients, although grade levels covered may be appropriate. This may result in scoring and interpretation difficulties. Special low SES norms a plus. Easy to administer and widely used test. Probably best for screening of basic achievement abilities and monitoring progress; may not provide enough specific diagnostic information for JTPA programs; therefore additional diagnostic measures may be required. Coverage: Reading - Adequate; Mathematics - Above Average; Written Communication - Limited (Spelling only); Verbal Communication - None; Problem-Solving - Very Limited (mathematics problems only).

Test: Tests of Achievement and Proficiency (TAP, 1985 Edition)

Publisher: Riverside Publishing Company
8420 Bryn Mawr Ave.
Chicago, Illinois 60631

Norms: Re-normed in 1985, Forms G & H. Grade levels 9-12, ages 15+. Special norms for lower SES groups available. Scores available: grade equivalent, national percentile rank, standard scores and normal curve equivalents.

Administration: Can be given by by teacher or other trained persons. Testing Guide available which includes administration instructions and guidelines for interpretation of results. Basic battery requires 160 minutes, complete battery requires 240 minutes. Hand or computerized scoring.

Cost: Approximately $4.00 per student.

Subtest Areas: Reading comprehension, mathematics, written expression, using sources of information. Areas within complete battery include social studies, science, listening test, writing test.

Comment: Norms are age appropriate for JTPA clients. Because of test requirements, some low functioning clients may have difficulty performing even at a basic level in some areas. Complete battery testing time is long. Special low SES norms are a plus. Easy to administer and becoming widely used in public schools. Probably best for screening of basic achievement abilities and to monitor progress, may not provide enough diagnostic information for JTPA programs, therefore additional diagnostic measures may be necessary. Coverage: Reading - adequate; Mathematics - adequate; Written Expression - adequate to above average if optional Writing Test is included; Verbal Communication - None; Problem Solving - very limited in Using Sources of Information (more academically related problem solving is assessed although some of it applies to practical problem solving).

Test: Cognitive Abilities Test

Publisher: The Riverside Publishing Company
8420 Bryn Mawr Ave.
Chicago, Illinois 60631

Norms: Revised and re-normed in 1985, Levels A-H. Covers Grades 3-12, ages 8+. Scores available: normal curve equivalents, standard age scores.

Administration: Can be given by teacher or other trained persons. Testing guide containing administration and interpretation information available. Test requires approximately 90 minutes. Hand and computerized scoring is possible.

Costs: Approximately $4.00 per student.

Subtest Areas: Verbal battery (verbal classification, sentence completion, verbal analogies), quantitative battery (quantitative relations, number series, equation building), nonverbal battery (figure classification, figure analogies, figure analysis)

Comment: Norms are age appropriate for JTPA clients. Easy to administer, more difficult to interpret. Probably best for diagnostic purposes, does not provide achievement level information. Coverage: Reading - none; Mathematics - very limited; Written Communication - none; problem-solving - limited.

Test: MULTISCORE

Publisher: The Riverside Publishing Company
8420 Bryn Mawr Ave.
Chicago, Illinois 60631

Norms: Available for grades 1 - adult. Criterion-referenced test which provide minimum proficiency criteria.

Administration: Can be given by teacher of other trained persons. Testing guide available which includes administration guidelines and scoring information. Testing time depends on number of objectives examined. Hand or computerized scoring available.

Cost: Depends on length of test, and number of clients being evaluated. A minimum number of tests must be ordered; each costs between $2.00 and $7.00 per student if ordered in large quantities (300-1000 minimum).

Subtest Areas: This is a customized criterion-referenced test which is developed according to user needs and specification. The system can provide tests of reading and language arts; mathematics; and science, social studies, and life skills.

Comment: Norms can typically be developed for a wide range of clients, some of which should be similar to JTPA clients. Requires user to choose the instructional objectives for testing which is a very technical and sophisticated process, but once done, can provide very efficient, and customized tests for many purposes. Very useful for screening and monitoring progress in a program, may be designed for diagnostic purposes in some areas. Coverage: reading - adequate; mathematics - adequate; written communication - limited; verbal communication - none; problem-solving - limited. Note that coverage in this case is dependent on best situation and appropriately designed assessment package.

Test: Gates-MacGinitie Reading Test

Publisher: The Riverside Publishing Company
8420 Bryn Mawr Ave.
Chicago, Illinois 60631

Norms: Normed in 1977, Forms A - F. Covers Grades 1-12, ages 6-adult. Scores available: grade equivalent, extended scale score, percentiles, normal curve equivalent, and stanines.

Administration: Can be given by teacher or other trained persons. Testing Guide provides administration and interpretation instructions. Basic testing time 55 minutes. Hand or computerized scoring is available.

Costs: Approximately $4.50 per student.

Subtest Areas: Reading vocabulary and reading comprehension.

Comment: Norms are age appropriate for JTPA clients. Easy to administer and widely used test. Probably best for screening of basic reading abilities and may be useful for diagnostic purposes also. Coverage: reading - above average; mathematics - none; written communication - none; verbal communication - none; problem-solving - None.

Test: Woodcock-Johnson Psycho-Educational Battery

Publisher: DLM Teaching Resources
P.O. Box 4000
One DLM Park
Allen, Texas 75002

Norms: Normed 1977. Grades K-12+, ages 3-80. Norms available by race, SES, sex and occupational status. Scores available: age and grade equivalents, percentiles, standard scores, difference scores, functioning levels.

Administration: Individually administered by trained examiner. Administration and interpretation manuals available. Testing time depends on number of tests used, ranges from 15 minutes to 4 hours or more.

Cost: Complete kit $165.00, plus about $1.00 per student.

Subtest Areas: Part One - Cognitive Ability Tests: Picture Vocabulary, Spatial Relations, Memory for Sentences, Visual-Auditory Learning, Blending, Quantitative Concepts, Visual Matching, Antonyms-Synonyms, Analysis-Synthesis, Numbers Reversed, Concept Formation, Analogies.

Part Two - Achievement Tests: Letter Word Identification, Word Attack, Passage Comprehension, Calculation, Applied Problems, Dictation, Proofing, Punctuation and Capitalization, Spelling, Usage, Science, Social Studies, Humanities.

Part Three - Interest Levels: Reading Interest, Math Interest, Written Language Interest, Physical Interest, Social Interest.

Part Four - Adaptive Behaviors (via interview): Gross Motor Skills, Fine Motor Skills, Social Interactions, Language Comprehension, Language Expression, Eating and Meal Preparation, Toileting, Dressing, Personal Self-Care, Domestic Skills, Time and Punctuality, Money and Value, Work Skills, Home/Community Orientation.

Comment: One of the best technically developed tests available which is also one of the most comprehensive in scope. Requires individual administration by trained examiner. Choice of tests and interpretation of results typically requires educational diagnostician, school or clinical psychologist for maximum information.

Subtests can be chosen for screening, monitoring, or diagnostic purposes. Strength is in diagnostic capability but requires extensive testing time. Coverage: reading - above average; mathematics - above average; written communication - limited; verbal communication - adequate in adaptive behavior domain; problem-solving - adequate.

Test: Test of Written Language (TOWL)

Publisher: Pro-Ed
5341 Industrial Oaks Blvd.
Austin, Texas 78735

Norms: Normed in 1983. Ages 7-19. Scores available: standard scores, percentile ranks, written language quotients.

Administration: Individually administered by trained examiner. Testing and interpretation guide available. Testing time is 40 minutes. Hand scoring.

Cost: Kit $55.00 plus about $1.00 per subject.

Subtest Areas: Vocabulary, thematic maturity, spelling, word usage, style, handwriting.

Comment: Norms are within age range for JTPA clients. Easy to administer but interpretation somewhat difficult and may require professional input. One of the more comprehensive tests of written communication abilities available. Useful for both screening, monitoring, and especially diagnostic purposes. Coverage: Reading - None; Mathematics - None; Written Communication - Above Average; Verbal Communication - None; Problem Solving - None.

Test: Gray Oral Reading Test (GOAT)

Publisher: Pro-Ed
5341 Industrial Oaks Blvd.
Austin, Texas 78735

Norms: Re-normed in 1986. Ages 7-17. Scores available: Standard scores and percentiles.

Administration: Individually administered by a trained examiner. Testing Guide available which includes administration instructions and interpretation information. Testing time is 10-20 minutes. Hand scoring required.

Cost: Kit approximately $70.00 plus about $1.25 per student.

Subtest Areas: Oral Reading and Oral Reading Comprehension.

Comment: Norms include the appropriate age range for JTPA clients. Giving and scoring test requires trained examiner who can code types of reading errors when they occur. One of most widely used oral reading tests. Useful for screening and diagnostic purposes. Coverage: reading - adequate; mathematics - none; written communication - none; verbal communication - none; problem-solving - none.

Test: Detroit Tests of Learning Aptitude

Publisher: Pro-Ed
5341 Industrial Oaks Blvd.
Austin, Texas 78735

Norms: Renormed in 1985. Ages 6-17. Scores available: standard scores and percentiles.

Administration: Individually administered by trained examiner. Testing and interpretation guide available. Testing times depend on which subtests given; times range between 15 minutes and 2+ hours. Hand or computerized scoring available.

Costs: Kit approximately $90.00 plus $1.00 per student.

Subtest Areas: Word opposites, sentence imitation, oral directions, word sequences, story construction, design reproduction, object sequences, symbolic relations, conceptual matching, word fragments, letter sequences.

Comment: Norms available for most JTPA age clients. Requires trained examiner and interpretation. Probably best for diagnostic purposes. Does not provide direct achievement measures. Coverage: reading - none; mathematics - none; written communication - none; verbal communication - none; problem-solving - limited.

Test: Wide Range Achievement Test-Revised (WRAT-R)

Publisher: Western Psychological Services
12031 Wilshire Blvd.
Los Angeles, CA. 90025

Norms: Revised and re-normed 1984. Ages 5-adult. Scores available: standard scores, percentiles, grade levels.

Administration: Can be given by teacher or other trained persons. Testing guide available which includes administration and limited interpretation guidelines. Test takes 15-30 minutes. Hand scoring.

Cost: Kit approximately $40.00 and $.50 per student.

Subtest Areas: Reading, spelling, arithmetic.

Comments: Norms include age range appropriate for JTPA clients. Probably the most easily administered, scored and widely used achievement screening test. Can also be used for monitoring purposes. Provides little diagnostic information expect to the skilled interpreter. Coverage: Reading - very limited; mathematics - limited; written communication - very limited; verbal communication - none; problem-solving - very limited (math problems only).

Test: Kaufman Test of Educational Achievement (K-TEA)

Publisher: American Guidance Service
Publishers' Building
P.O. Box 99
Circle Pines, MN 55014-1796

Norms: Normed in 1985. Grades 1-12, ages 6 - 18. Special subgroup norms available. Scores available: age-based standard scores, percentiles, normal curve equivalents, stanines.

Administration: Individually administered by trained examiner. Testing guide available which includes administration instructions and interpretation guidelines. Brief Form takes between 20-30 minutes, while more Comprehensive Form takes up to 75 minutes. Hand scoring.

Cost: Kit approximately $110.00 and $1.00 per student.

Subtest Areas: Brief Form: mathematics, reading, spelling. Comprehensive Form: mathematics application, reading decoding, spelling, reading comprehension, mathematics computation.

Comment: Norms are within age or grade ranges of most JTPA clients. Special low SES norms a plus. Easy to administer, but requires some expertise for interpretation. Fairly new test. Probably best for screening, monitoring, and some diagnostic purposes if comprehensive form used. Coverage: reading - adequate; mathematics - adequate; written communication - very limited (only spelling); verbal communication - none; problem-solving - very limited (math-related only).

Test: Adult Basic Learning Examination (ABLE - 2nd Edition)

Publisher: The Psychological Corporation
555 Academic Court
San Antonio, Tx 78204-0952

Norms: Revised and normed 1986. Adults. Special norms for adult clients with various educational backgrounds. Scores available: scaled scores, percentiles, stanines, grade equivalents.

Administration: Can be given by teacher or other trained examiner. Scoring and interpretation manual available. Testing time depends on educational background: 1-4 years of schooling requires 130 minutes, 5+ years of schooling requires 175 minutes. Hand and computer scoring is possible.

Cost: Approximately $1.50 per student.

Subtest Areas: Vocabulary, reading comprehension, spelling, number operations, problem-solving, applied grammar, and capitalization/punctuation subtests.

Comment: Norms are some of the best available for JTPA clients. Easy to administer and score, was designed for adult education purposes. Best considered a screening and monitoring test, diagnostic information available to trained interpreter. Coverage: reading - adequate; mathematics - adequate; written communication - limited; verbal communication - none; problem-solving - very limited (mathematics area only). Note: older edition of this test does not qualify for these comments.

Test: Tests of Adult Basic Education (TABE) (Forms 5 and 6)

Publisher: McGraw/Hill
Publishers Test Service
2500 Garden Road
Monterey, CA 93940-5380

Norms: Revised and renormed, 1987. Grade 2-12, adults. Special norms for adult students. Scores available: grade equivalents, percentiles.

Administration: Can be given by teacher or other trained persons. Testing guide with administration instructions and interpretation information. Survey form takes about 60-100 minutes, the complete battery takes about 200 minutes. Hand or computerized scoring is possible.

Cost: Approximately $2.00 per student.

Subtest Areas: Reading (vocabulary, comprehension), language skills (mechanics, expression), and mathematics (computation, concepts and application), spelling.

Comment: Probably one of the most widely used tests for adult students. Norms appropriate for JTPA clients. Is best used for screening and monitoring, but can provide some diagnostic information (instructional levels) from qualified interpreter. Was originally derived from California Achievement Test, so provides similar information. Coverage: reading - adequate; Mathematics - adequate; written communication - very limited (spelling only); verbal communication - none; problem-solving - none.

Test: USES Basic Occupational Literacy Test (BOLT)

Publisher: U.S. Government Printing Office
Must get state employment security office approval to obtain and use.

Norms: Normed 1974. BOLT has four levels, measuring grades 1-12. Special norms for educationally disabled adults available. Scores available: scaled scores and percentiles.

Administration: Instruction manual available. Can only be given by approved users. Testing takes 90-130 minutes. Hand or computerized scoring is available.

Cost: Varies by state.

Subtest Areas: Reading vocabulary, reading comprehension, arithmetic computation, arithmetic reasoning.

Comment: Required ability levels may be too high for some JTPA clients. Norms are SES appropriate, but other characteristics may not match JTPA clients. Best used for screening and monitoring, can provide some limited diagnostic information. Coverage: reading - adequate; mathematics - adequate; written communication - none; verbal communication - none; problem-solving - very limited (math area only).

Test: Woodcock Reading Mastery Tests - Revised

Publisher: American Guidance Service
Publishers' Building
P.O. Box 99
Circle Pines, MN 55014-1796

Norms: Normed 1985. Grades K-college, ages up to 70. Scores available: standard scores, percentiles, age and grade equivalents, NCEs.

Administration: Can be given by teacher or trained diagnostician. Testing guide available. Testing times: Brief scale - 15 minutes, complete test - up to 90 minutes. Hand scoring.

Cost: Kit approximately $70.00 plus $1.00 per student.

Subtest Areas: Form G - visual auditory learning, letter identification, word identification, word attack, word comprehension, passage comprehension. Form H: word identification, word attack, word comprehension, passage comprehension.

Comment: Norms are extensive. Interpretation requires trained person. Can be used for screening, probably best for diagnostics. Coverage: reading - above average; mathematics - none; written communication - none; verbal communication - none; problem-solving - none.

Test: Key Math Diagnostic Test

Publisher: American Guidance Service
Publishers' Building
P.O. Box 99
Circle Pines, MN 55014-1796

Norms: Normed in 1976. Grades 2-6. Scores available: Normal Curve Equivalents (NCEs), grade equivalents, percentile.

Administration: Can be given by teacher or other trained person. Testing guide available. Testing time is 30-40 minutes. Hand and computerized scoring available.

Cost: Kit approximately $55.00 plus $.50 per student.

Subtest Areas: Numeration, fractions, geometry and symbols, addition, subtraction, multiplication, division, mental computation, numerical reasoning, word problems, missing elements, money, measurement, time.

Comment: Norms may not be appropriate for JTPA clients due to age. Probably best used for diagnostic purposes, but screening information also available. Coverage: reading - none; mathematics - above average; written communication - none; verbal communication - none; problem-solving - very limited (math only).

Test: Peabody Individual Achievement Test

Publisher: American Guidance Service
Publishers' Building
P.O. Box 99
Circle Pines, MN 55014-1796

Norms: Normed 1970. Grades K-12 (adults). Scores available: grade and age equivalents, standard scores, percentiles.

Administration: Can be given by teacher or other trained persons. Testing Guide available. Testing time is 30-50 minutes. Hand scoring.

Cost: Kit approximately $75.00 plus $.50 per student.

Subtest Areas: Mathematics, reading recognition, reading comprehension, spelling, general information.

Comment: Norms are within age and grade range of JTPA clients. Easy to administer. Probably best for screening and monitoring purposes. Coverage: reading - adequate; mathematics - adequate; written communication - very limited (spelling only); verbal communication - none; problem-solving - very limited (math only).

Test: CASAS Adult Life Skills Pre-Employment Tests

Publisher: Comprehensive Adult Student Assessment System and the San Diego Community College District Foundation
2725 Congress Street, Suite 1-M
San Diego, CA 92110

Norms: Criterion- and competency-referenced tests which provide functional proficiency criteria for competency-based employment-related programs.

Administration: May be group or individually administered by trained persons. CASAS provides training on administering and interpreting the tests.

Cost: Depends on number and type of types administered.

Subtest Areas: Employability Competency System Appraisal for initial identification of basic reading and math functional skill levels in an employability context.

Survey Achievement Tests:
- at three levels (A, B, and C) for monitoring progress in reading.
- at two levels (A and B) for monitoring progress in math.
- at three levels (A, B, and C) for monitoring progress in listening comprehension.

All areas are tested in an employability context and alternate forms are available for each level.

Certification Tests in an employability context for two levels (B and C) in reading and math for determining level or program completion.

Comment: The CASAS assessment design includes a bank of more than 4000 items that have been extensively field tested throughout California and other states over an eight-year period. Each item is designed to measure a specific competency statement but also on a continuum of difficulty as he/she progresses through the program. The underlying common achievement scale based on Item Response Theory allows for better articulation among programs and levels. Individual Achievement can be monitored, as well as group progress because all items have been calibrated on the same scale. The tests are appropriate for native and non-native English speakers functioning from minimal through high school entry level skill levels.

APPENDIX B

JTPA SURVEY CONDUCTED BY THE CENTER FOR REMEDIATION DESIGN WITH BRANDEIS UNIVERSITY

SURVEY FORM

Part 1: Basic Skills Remediation in JTPA Youth Programs

1. Do you provide basic skills remediation for JTPA youth?
   _____ yes, summer only (IIB)
   _____ yes, school year only
   _____ yes, both summer and school-year
   _____ no

2. Who is served in your program(s)? (Check all applicable)
   _____ in-school youth
   _____ dropouts
   _____ high school graduates

3. Describe your program's instructional technique. (Check all applicable)
   _____group instruction
   _____individual/self-paced
   _____competency-based
   _____computers are used as teaching tools
   _____instruction is specifically tied to work experience
   _____instruction is specifically tied to skills training

4. How would you rate the results of your program?
   _____Excellent _____Good _____Fair _____Poor

5. How is your remediation program funded?
   _____ JTPA 8%
   _____ JTPA IIA
   _____ JTPA IIB
   _____ other (please be specific)

6. Is your remediation program linked to a JTPA youth competency system? If yes: are competency gains measured by grade level scores? Functional skill gains? GED test? Other:_____

7. What do you see as the three biggest problems in providing remediation to youth in your programs? (Topics to be covered in the paper).

Part 2: JTPA Assessment Strategies: Identifying Issues and Instruments

8. Do you provide formal testing for youth in remediation in IIA? in IIB? (standardized)

9. If you administer a formal test(s) what do you use in IIA? in IIB? (list all that apply)

10. How do you use assessment information?

test	to sort	to diagnose	for progress checks	credentialing

11. What other assessment strategies besides tests do you use? Intake interview? Performance reviews (behavior observation)? Product development? Other?

12. Do you use information from other sources? If yes, what tests? What sources? (i.e., schools)

SURVEY METHODOLOGY

The survey was conducted during August, 1987. The summary that follows reports upon 150 programs out of an originally randomly selected sample of 205. (This sample was developed by taking every third SDA on an alphabetized list of approximately 610 SDA administrative entities.) Appendix B shows, state by state, the distribution of the sample and the number of individuals contacted. If no bias was introduced by the sample not being completed, the sample size is probably adequate for the purpose intended (with an error of not more than 8% at the 95% level of confidence). There does not seem to be any obvious variation of responses between states. Many states had only one respondent and comparison is therefore undependable. The one broad comment which can be made is that the variation of response within states seems to depend mainly upon the number of respondents within the state.

The interviewers asked to speak to the person in charge of the SDA's youth program. First contacts were not usually well informed about the programs in operation. Further referrals (often as many as seven) proved to be of greater help and were more enthusiastic about programming efforts. As a rule, JTPA program operators tended to have more information than the SDA or the PIC contacts.

SURVEY FINDINGS

Question One: Provision of Basic Skills Remediation

69.3% of the programs sampled provide basic skills remediation both in summer and during the school year, 28% during the summer only and 2% during the school year only. (One response was not available.)

Question Two: Who Do The Programs Serve?

Programs typically serve youth who are still in school, together with others no longer in school - this combination represents 92% of the sample. Other target groups for service were all encompassing.

- 68% of the sample had programs which served in-school youth, dropouts and high school graduates;
- 16.6% in-school youth and dropouts;
- 2% in-school youth and high school graduates; and
- 7.3% in-school youth only.

One response was not available. The remaining 2.6% of the sample offered programs to dropouts and high school graduates only.

Question Three: Instructional Techniques

Most programs used a variety of instructional techniques varied by program and client need.

- 70.7% used computers as teaching tools;
- 74% used competency based techniques;
- 57.3% tied instruction to work experience;
- 53.3% tied instruction to skills training;
- 73.3% used individual/self paced techniques; and
- 73.3% used group instruction.

The most common combination of techniques was to use all of them; this was the case for 24.7% of respondents. The next most common combination of techniques was to use group instruction, individual/self paced instruction, competency based instruction and computers as teaching tools; this combination was used by 9.3% of respondents.

Question Four: Program Results

Perhaps predictably, respondents rated their program results very highly.

- 27.6% claimed to have excellent results;
- 57.7% reported good results;

- 5.3% reported fair results;
- One respondent (0.7% of the sample) reported poor results; and
- 8.7% of the sample gave no response.

Question Five: Program Funding

Funds for the programs most typically came from JTPA exclusively and were usually derived from a combination of sources.

- 2.7% were funded from JTPA 8% only;
- 8.7% were funded from JTPA IIA only;
- 28.7% were funded from JTPA IIB only;
- 29.3% were funded from a combination that included JTPA 8%;
- 58.7% were funded from a combination that included JTPA IIA;
- 86% were funded from a combination that included JTPA IIB; and
- The most common combination was that of JTPA IIA and JTPA IIB which was used by 30.7% of respondents.

Question Six: Linkage to Youth Competency System

85% of the programs were linked to a JTPA youth competency system. Of these, the most common procedures for defining outcome or attainment were:

- Grade level scores - 24.3%;
- Functional skill gains - 21.3%; and
- A combination of grade level scores and GED test - 13.3%.

No other option was used by more than 10% of the sample.

Question Seven: Problems in Providing Remediation

Most respondents mentioned more than one problem in providing remediation. The two most often mentioned problems were "motivation and type of incentive programs" and "remediation problems and attendance" (32.7% of the respondents mentioned these two problems; 14% mentioned "motivation and type of incentive programs" as the only problem). Other significant problems:

- 16.7% mentioned "role clarification of JTPA vs. school responsibilities for youth";
- 15.3% mentioned "lack of cooperation from school system";
- 13.3% mentioned recruitment;
- 12% mentioned rural county problems; and
- 10.7% mentioned transportation.

No other problem was mentioned by more than 10% of respondents.

Question Eight: Formal Testing for Remediation

92% of the programs provided formal testing for youth in remediation.

Question Nine: Tests Used

Of those JTPA programs which administer formal tests themselves the following emerged as the most commonly used:

- TABE is used by 39.3% of programs;
- CAT is used by 22.7% of programs;
- WRAT is used by 16.7% of programs;
- ABLE was used by 9.3% of programs; and
- 7.3% of tests used were self-made.

None of the other tests mentioned was used by more than four respondents (2.7% of the sample).

Question Ten: Use of Assessment Information

Assessment information was used for a combination of purposes by most programs.

- 34.7% used it to sort youth into groups (appraisal);
- 68.7% used it to diagnose where learning should begin within a defined level;
- 30.7% used it for progress checks (benchmarking); and
- 66% used it for certifying attainment.

Question Eleven: Other Assessment Strategies

The most common additional assessment strategy used was the intake interview, which was used by 44.7% of respondents. None of the other strategies, or combination of strategies, was used by more than 10% of respondents.

Question Twelve: Information from Other Sources

Information from schools was the only other commonly mentioned source of information; 95.3% of respondents mentioned school as an information source. No other source was mentioned by more than one program.

SURVEY PARTICIPANTS

State	Number of SDAs in the Sample	Number of Persons Contacted in Each State
Alaska	1	1
Alabama	1	1
Arkansas	2	3
Arizona	5	6
California	15	17
Colorado	3	4
Connecticut	3	4
Florida	6	8
Georgia	4	6
Hawaii	1	2
Iowa	5	5
Idaho	0	2
Illinois	3	8
Indiana	6	6
Kansas	2	2
Kentucky	2	3
Louisiana	3	6
Massachusetts	5	5
Maryland	4	4
Michigan	6	8
Minnesota	2	6
Missouri	2	5
Montana	1	1
North Carolina	9	9
Nebraska	1	1
New Hampshire	0	1
New Jersey	6	6
New Mexico	0	1
New York	1	11
Ohio	6	10
Oklahoma	1	4
Oregon	2	2
Pennsylvania	8	10
Puerto Rico	1	1
Rhode Island	1	1
Tennessee	2	5
Texas	5	11
Utah	3	3
Virginia	4	4
Vermont	1	1
Washington	3	4
Wisconsin	3	6
West Virginia	0	1